

AMERICAN V 510 '35
. 002
JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

E. T. BELL
CALIFORNIA INSTITUTE OF TECHNOLOGY

E. W. CHITTENDEN
UNIVERSITY OF IOWA

ABRAHAM COHEN
THE JOHNS HOPKINS UNIVERSITY

G. C. EVANS
UNIVERSITY OF CALIFORNIA

F. D. MURNAGHAN
THE JOHNS HOPKINS UNIVERSITY

WITH THE COÖPERATION OF

FRANK MORLEY
HARRY BATEMAN
W. A. MANNING

MARSTON MORSE
J. R. KLINE
E. P. LANE
HARRY LEVY

ALONZO CHURCH
L. R. FORD
OSCAR ZARISKI

PUBLISHED UNDER THE JOINT AUSPICES OF
THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

Cuk - 1402329-52. per 14.71

VOLUME LVII
1935

THE JOHNS HOPKINS PRESS
BALTIMORE, MARYLAND
U. S. A.

P14900

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

INDEX.

	PAGE
ARBER, S. F. and ZARISKI, OSCAR. Reducible exceptional curves of the first kind,	119
BASOCO, M. A. and BELL, E. T. Further theta expansions useful in arithmetic,	1
BELL, E. T. See Basoco and Bell, page 1.	
BELL, E. T. Doubly periodic functions of the second kind and the arithmetical form $xy + zw$,	245
BELL, E. T. Arithmetical theorems on Lucas functions and Tchebycheff polynomials,	781
BERNSTEIN, B. A. On finite Boolean algebras,	733
BERRY, ANDREW C. Linear metric Lebesgue spaces,	743
BIRKHOFF, GARRETT. Abstract linear dependence and lattices,	800
BRAHANA, H. R. Metabelian groups and pencils of bilinear forms, . .	645
BRAUER, RICHARD and WEYL, HERMANN. Spinors in n dimensions, . .	425
BROWN, ARTHUR B. Isolated critical points,	389
BLUMENTHAL, LEONARD M. Concerning spherical spaces,	51
CHANLER, JOSEPHINE H. See Coble and Chanler, page 183.	
CHHO, SI-PING. Singularities of analytic vector functions,	294
COBLE, ARTHUR B. and CHANLER, JOSEPHINE H. The geometry of the Weddle manifold W_2 ,	183
CRAIG, H. V. On a generalized tangent vector,	457
DE DUFFAHEL, MAURICE. Gegenbauer functions with a negative index, .	918
DICKSON, L. E. Cyclotomy, higher congruences, and Waring's problem, .	391
DICKSON, L. E. Cyclotomy, higher congruences, and Waring's problem II, .	463
EVANS, GRIFFITH C. Correction and addition to "complements of potential theory,"	623
FARRELL, O. J. On the expansion of harmonic functions in series of harmonic polynomials belonging to a simply connected region, . .	883
FLEXNER, WILLIAM W. The intersection of chains on a topological manifold,	309
GAGE, WALTER H. Addition theorems for the doubly periodic functions of the second kind,	342
GARVER, RAYMOND. A definition of group by means of three postulates, .	276

	PAGE
HADLOCK, E. H. On the progressions associated with a ternary quadratic form,	267
HAVILAND, E. K. On the inversion formula for Fourier-Stieltjes transforms in more than one dimension,	94
HAVILAND, E. K. On the inversion formula for Fourier-Stieltjes transforms in more than one dimension. II,	382
HAVILAND, E. K. On the momentum problem for distribution functions in more than one dimension,	562
HAVILAND, E. K. A note on a property of Fourier-Stieltjes transforms in more than one dimension,	569
HEDLUND, GUSTAV A. A metrically transitive group defined by the modular group,	668
HINRICHSSEN, J. J. L. Note on potential theory in n -space,	921
HOEL, PAUL G. Certain problems in the theory of closest approximation,	891
HUGHES, OLIVE MARGARET. A certain mixed linear integral equation,	861
KERSHNER, RICHARD and WINTNER, AUREL. On symmetric Bernoulli convolutions,	541
KERSHNER, RICHARD. Determination of a van der Corput-Landau absolute constant,	840
KLEENE, S. C. A theory of positive integers in formal logic. Part I,	153
KLEENE, S. C. A theory of positive integers in formal logic. Part II,	219
KRALL, H. L. On some asymptotic relations for the characteristic values of the elliptic differential equations,	907
LOWAN, ARNOLD N. On the cooling of the earth,	174
LUNN, A. C. See Senior and Lunn, page 254.	
MACLANE, SAUNDERS. Some unique separation theorems for graphs,	805
MACQUEEN, M. L. The conjugate chord quadrics of a curve on a surface,	771
MCCOY, NEAL H. On the rational canonical form of a function of a matrix,	491
MCEWEN, W. H. Polynomials of best approximation associated with certain problems in two dimensions,	367
MILLER, G. A. Groups containing five and only five squares,	615
MORREY, CHARLES B., JR. The topology of (path) surfaces,	17
MORREY, CHARLES B., JR. An analytic characterization of surfaces of finite Lebesgue area. Part I,	692
MOURSUND, A. F. On summation of derived series of the conjugate Fourier series,	854

	PAGE
MUSSELMAN, J. R. On certain types of hexagons,	503
PALL, GORDON. Binary quadratic discriminants differing by square factors,	789
PETERS, R. M. Parallelism and equidistance in Riemannian geometry,	103
PRICE, G. BAILEY. The generalized double pendulum,	928
RADÓ, TIBOR. The isoperimetric inequality on the sphere;	765
REID, WILLIAM T. Discontinuous solutions in the non-parametric problem of Mayer in the calculus of variations,	69
REID, WILLIAM T. The theory of the second variation for the non-parametric problem of Bolza,	573
SENIOR, J. K. and LUNN, A. C. Determination of the groups of orders 162-215 omitting order 192,	254
SHEFFER, I. M. Concerning some methods of best approximation, and a theorem of Birkhoff,	587
STONE, M. H. Postulates for Boolean algebras and generalized Boolean algebras,	703
SUCKAU, J. W. THEODORE. On uniform convergence,	549
SYNGE, J. L. Some intrinsic and derived vectors in a Kawaguchi space,	679
TARTLER, A. On a certain class of orthogonal polynomials,	627
THOMPSON, WILLIAM R. On the theory of apportionment,	450
VAN KAMPEN, E. R. The topological transformations of a simple closed curve into itself,	142
VAN KAMPEN, E. R. The structure of a compact connected group,	301
WARD, LEWIS E. A third-order irregular boundary value problem and the associated series,	345
WARD, MORGAN. A determination of all possible systems of strict implication,	261
WEYL, HERMANN. See Brauer and Weyl, page 425.	
WHEELER, ANNA PELL. Spectral theory for a certain class of non-symmetric completely continuous matrices,	847
WHITNEY, HASSLER. On the abstract properties of linear dependence,	509
WHYBURN, G. T. Concerning continua of finite degree and local separating points,	11
WHYBURN, G. T. Regular convergence and monotone transformations,	902
WILLIAMSON, J. The simultaneous reduction of two matrices to triangle form,	281

	PAGE
WILLIAMSON, J. The equivalence of non-singular pencils of Hermitian matrices in an arbitrary field,	475
WILSON, W. A. On certain types of continuous transformations of metric spaces,	62
WILSON, W. A. On the imbedding of metric sets in euclidean space,	322
WINTNER, AUREL. A note on the distribution of the zeros of the zeta-function,	101
WINTNER, AUREL. On the differentiation of infinite convolutions,	363
WINTNER, AUREL. On the asymptotic distribution of the remainder term of the prime-number theorem,	534
WINTNER, AUREL. On the exact value of the bound for the regularity of solutions of ordinary differential equations,	539
WINTNER, AUREL. See Kershner and Wintner, page 541.	
WINTNER, AUREL. Gaussian distributions and convergent infinite convolutions,	821
WINTNER, AUREL. On convergent Poisson convolutions,	827
WINTNER, AUREL. A note on the convergence of infinite convolutions,	839
WONG, B. C. On a rational surface of order 12 in 9-space and its projections,	112
WYLIE, C. R., JR. Space curves belonging to a non-special linear line complex,	937
ZARISKI, OSCAR. See Barber and Zariski, page 119.	
ZIPPIN, LEO. On semicompact spaces,	327

FURTHER THETA EXPANSIONS USEFUL IN ARITHMETIC.

By M. A. BASOCO AND E. T. BELL.

1. In a paper with a similar title † several such expansions for functions of one variable are given, all the expansions in the paper being deduced from those of the so-called doubly periodic functions of the second kind,

$$\vartheta'_{\alpha\beta\gamma} \frac{\vartheta_{\alpha}(x+y)}{\vartheta_{\beta}(x)\vartheta_{\gamma}(y)},$$

where the triple index $\alpha\beta\gamma$ has the values,

001,	010,	023,	032,
100,	111,	122,	133,
203,	212,	221,	230,
302,	313,	320,	331.

The complete set of these sixteen expansions (in trigonometric series) seems to have been first obtained by Hermite, although some appear implicitly in Jacobi's memoir on the rotation of a rigid body. Another method, due to Kronecker, of deriving the sixteen expansions is well known. Misprints and other inaccuracies in the final forms have been carried over to some of the standard treatises; for an exact list the reader is referred to previous papers, of which only one need be cited.‡

For the full and efficient use of the method of paraphrase (paper cited in preceding footnote), it is necessary to have the trigonometric expansions of the remaining forty-eight functions

$$K_{abc} \frac{\vartheta_a(x+y)}{\vartheta_b(x)\vartheta_c(y)},$$

where K_{abc} is a theta constant (independent of x, y), which is to be determined for a given abc other than one of the sixteen values belonging to Hermite's functions. All sixty-four expansions when used in paraphrase have the desirable effect of introducing cross-product terms into the quadratic forms appearing in the partitions. This has been amply illustrated in former papers for the sixteen quoted values of abc .

† E. T. Bell, *Messenger of Mathematics*, vol. 54 (1924), pp. 166-176.

‡ E. T. Bell, *Transactions of the American Mathematical Society*, vol. 22 (1921), pp. 1-30 and 198-219.

There remain, then, forty-eight fundamental expansions to be determined. The constant K_{abc} in a given instance, it will be noticed, is subject to arbitrary choice; a change in the constants, of course, radically alters the nature of the arithmetical functions appearing in the expansions. For choosing the 48 constants K_{abc} , there are two apparently reasonable clues—one choice being suggested by a discussion due to Krause,[†] the other being indicated if we require products of the type

$$K_{a\beta\gamma} \frac{\vartheta_a(x+y)}{\vartheta_\beta(x)\vartheta_\gamma(y)} \cdot K_{abc} \frac{\vartheta_a(x-y)}{\vartheta_b(x)\vartheta_c(y)}$$

to be bilinear in the squares of the twelve elliptic functions Asn, Bcn, Cns, \dots of Glaisher, where A, B, C, \dots are the classical constants $2K/\pi$, etc. The reason for the last requirement is obvious from its interpretation in terms of paraphrases. Both methods indicate the same constants. Following the suggestion of Krause, Professor D. A. F. Robinson in a forthcoming paper in the *Transactions of the Royal Society of Canada* has given the first (so far as we know) determination of these important series. If an alternative set of 48 expansions appears by the second method, we shall have an immediate source of arithmetical theorems on comparing with Robinson's. Further, either set can be used in any application of paraphrasing.

In neither set of 48 expansions has it been possible to refer only to the divisors of a simple integer, as is the case for Hermite's 16. Instead (in the present method) we need all solutions in positive integers d, δ, d', δ' of equations of the form $n = d\delta + d'\delta', n = d\delta + 2d'\delta'$, etc.

Following a method suggested elsewhere[‡] we have calculated the 48 expansions very simply from the 16. This may be done in two ways, the partitions involved being, in most cases, different, so that a comparison of the two results yields, at once, a series of arithmetical results. In this paper we record the arithmetical forms of the expansions (both sets), and a sufficient indication (in § 8) of the means by which they can be recalculated. Although no applications are included, it may be mentioned that the complete set of 64 gives an endless variety of paraphrases involving functions of 2, 4, 6, 8 \dots variables, the functions being both *complete* and *incomplete*.

The 48 functions correspond to the following values of the triple index abc :

[†] Krause, *Mathematische Annalen*, vol. 30 (1887), pp. 425-436 and pp. 516-534.

[‡] M. A. Basoco, *American Journal of Mathematics*, vol. 54 (April, 1932), pp. 242-252 (p. 242, footnote).

110,	213,	011,	312,
120,	223,	021,	322,
130,	233,	031,	332,
<hr/>			
000,	303,	202,	101*,
020,	323*,	222,	121,
030,	333,	232*,	131,
<hr/>			
300,	003*,	102*,	201,
310,	013*,	112*,	211,
330*,	033,	132,	231*,
<hr/>			
103*,	200,	301*,	002*,
113*,	210*,	311,	012*,
123*,	220*,	321*,	022.
<hr/>			

A star (*) indicates that the function with the starred index can be obtained by interchanging x, y in some function already in the table, and hence its expansion need not be separately calculated (except as a check). The reason for the particular grouping into sets of 12 above will appear if the expansions are verified by the method suggested in § 8. All reductions will be omitted and only the final results stated.

2. The m, n, d, δ, t, τ , notation for integers > 0 of former papers is used. Letters with suffixes or accents denote integers of the same respective kinds as those without suffixes or accents; m, τ are odd; n, d, δ, t are unrestricted (odd or even). The first Σ in a given expansion refers to *all* m or n (as defined above) in the exponent of q . The second Σ occurs in the coefficient of the general power of q , and refers to *all* solutions, for m or n fixed, of the indicated equations (or partitions) at the head of a particular set of expansions.

For easy reference we state all the partitions here:

- (I) $m = 2n_1 + m_2; \quad m = t\tau, \quad n_1 = t_1\tau_1, \quad m_2 = t_2\tau_2.$
 (II) $n = n_1 + n_2; \quad n = t\tau, \quad n_1 = t_1\tau_1, \quad n_2 = t_2\tau_2.$
 (III) $\begin{cases} m = 2n_1 + n_2; & m = t\tau, & n_1 = d_1\delta_1, & n_2 = t_2\tau_2, \\ 2n = 2n'_1 + n'_2; & 2n = t'\tau', & n = d\delta, & n'_1 = d'_1\delta'_1, & n'_2 = t'_2\tau'_2. \end{cases}$
 (IV) $m = m_1 + 4n_2; \quad m = t\tau, \quad m_1 = t_1\tau_1, \quad n_2 = d_2\delta_2.$
 (V) $2n = 2n_1 + 2n_2; \quad n = d\delta, \quad n_1 = d_1\delta_1, \quad n_2 = d_2\delta_2.$

We shall write

$$\phi_{abc}(x, y) \equiv K_{abc} \psi_{abc}(x, y), \quad \psi_{abc}(x, y) \equiv \frac{\partial_a(x+y)}{\partial_b(x)\partial_c(y)},$$

$(-1/h) \equiv (-1)^{(h-1)/2}$ if h is an odd integer. Since

$$\phi_{abc}(x, y) = \phi_{acb}(y, x),$$

it is sufficient to state the expansion of one of these functions. The constants K_{abc} are to be specified with the expansions.

3. There are 20 expansions in which the partition is (I) of § 2. All are of the form

$$\phi_{abc}(x, y) = 4 \sum_{(m)} \bar{q}^{m/2} (\sum F_{abc}(x, y; m)),$$

the first Σ referring to all m (in the notation explained in § 2), the second to all integers defined by the partition (I) for m fixed; namely, $\Sigma F_{abc}(x, y; m)$ is the coefficient of $q^{m/2}$. To state the expansions concisely it is sufficient to tabulate K_{abc} and $F_{abc}(x, y; m)$ for the 20 indices abc concerned. Attending to the remark at the end of § 2, we need only the following 13.

	abc	K_{abc} (Partition (I))
(1)	110	$\partial'_1 \partial_2 \partial_3$
(2)	120	$\partial'_1 \partial_0 \partial_2$
(3)	130	$\partial'_1 \partial_0 \partial_3$
(4)	213	$\partial'_1 \partial_2 \partial_3$
(5)	223	$\partial'_1 \partial_0 \partial_2$
(6)	233	$\partial'_1 \partial_0 \partial_3$
(7)	000	$\partial'_1 \partial_2 \partial_3$
(8)	030	$\partial'_1 \partial_0 \partial_2$
(9)	303	$\partial'_1 \partial_2 \partial_3$
(10)	333	$\partial'_1 \partial_0 \partial_2$
(11)	300	$\partial'_1 \partial_0 \partial_2$
(12)	033	$\partial'_1 \partial_2 \partial_3$
(13)	200	$\partial'_1 \partial_0 \partial_3$

The $F_{abc}(x, y; m)$ are numbered correspondingly and are as follows:

- (1) $\sin(tx + \tau y) \csc x - 2[\cos\{(\tau_1 + t_2)x + \tau_2 y\} - \cos\{(\tau_1 - t_2)x - \tau_2 y\}].$
- (2) $\sin(tx + \tau y) \sec x$
 $+ 2(-1)^m (-1/\tau_1) [\sin\{(\tau_1 + t_2)x + \tau_2 y\} - \sin\{(\tau_1 - t_2)x - \tau_2 y\}].$
- (3) $\sin(tx + \tau y)$
 $+ 2(-1)^m (-1/\tau_1) [\sin\{(2t_1 + t_2)x + \tau_2 y\} - \sin\{(2t_1 - t_2)x - \tau_2 y\}].$

- (4) $(-1/\tau) \cos(tx + \tau y) \csc x$
 $+ 2(-1/\tau_2) [\sin\{(\tau_1 + t_2)x + \tau_2 y\} + \sin\{(\tau_1 - t_2)x - \tau_2 y\}].$
- (5) $(-1/\tau) \cos(tx + \tau y) \sec x$
 $+ 2(-1)^{n_1} (-1/\tau_1 \tau_2) [\cos\{(\tau_1 + t_2)x + \tau_2 y\} + \cos\{(\tau_1 - t_2)x - \tau_2 y\}].$
- (6) $(-1/\tau) \cos(tx + \tau y)$
 $+ 2(-1)^{n_1} (-1/\tau_1 \tau_2) [\cos\{(2t_1 + t_2)x + \tau_2 y\} + \cos\{(2t_1 - t_2)x - \tau_2 y\}].$
- (7) $\sin tx \csc x - 2[\cos\{(t_2 + \tau_1)x + 2t_1 y\} - \cos\{(t_2 - \tau_1)x - 2t_1 y\}].$
- (8) $(-1/m) \sin tx \csc x$
 $- 2(-1/m_2) [\cos\{(t_2 + \tau_1)x + 2t_1 y\} - \cos\{(t_2 - \tau_1)x - 2t_1 y\}].$
- (9) $\sin tx \csc x$
 $- 2(-1)^{n_1} [\cos\{(t_2 + \tau_1)x + 2t_1 y\} - \cos\{(t_2 - \tau_1)x - 2t_1 y\}].$
- (10) $(-1/m) \sin tx \csc x$
 $- 2(-1)^{n_1} (-1/m_2) [\cos\{(t_2 + \tau_1)x + 2t_1 y\} - \cos\{(t_2 - \tau_1)x - 2t_1 y\}].$
- (11) $(-1/\tau) \cos tx \sec x$
 $+ 2(-1/\tau_1 \tau_2) [\cos\{(t_2 + \tau_1)x + 2t_1 y\} + \cos\{(t_2 - \tau_1)x - 2t_1 y\}].$
- (12) $(-1/t) \cos tx \sec x$
 $+ 2(-1)^{n_1} (-1/\tau_1 t_2) [\cos\{(t_2 + \tau_1)x + 2t_1 y\} + \cos\{(t_2 - \tau_1)x - 2t_1 y\}].$
- (13) $(-1/\tau) \cos(\tau x + ty)$
 $+ 2(-1/\tau_1 \tau_2) [\cos\{(2t_1 + \tau_2)x + t_2 y\} + \cos\{(2t_1 - \tau_2)x - t_2 y\}].$

4. There are twelve expansions in which the partition is (II). All are of the form

$$\phi_{abc}(x, y) = T_{abc}(x, y) + 4 \sum_{(n)} q^n (\sum F_{abc}(x, y; n)),$$

in which $T_{abc}(x, y)$ is a trigonometric term independent of the partition, and the rest of the notation is obvious from § 3. As in § 3 it is sufficient to tabulate here only 8.

	abc	K_{abc}	$T_{abc}(x, y)$ (Partition (II))
(1)	011	$\partial'_1 \partial_2 \partial_3$	$\csc x \csc y$
(2)	021	$\partial'_1 \partial_0 \partial_2$	$\sec x \csc y$
(3)	031	$\partial'_1 \partial_0 \partial_3$	$\csc y$
(4)	312	$\partial'_1 \partial_2 \partial_3$	$\csc x \sec y$
(5)	322	$\partial'_1 \partial_0 \partial_2$	$\sec x \sec y$
(6)	332	$\partial'_1 \partial_0 \partial_3$	$\sec y$
(7)	311	$\partial'_1 \partial_0 \partial_2$	$\csc x \csc y$
(8)	022	$\partial'_1 \partial_2 \partial_3$	$\sec x \sec y.$

The corresponding F 's follow.

- (1) $\sin(2tx + \tau y) \csc x + \sin \tau x \csc y$
 $- 2[\cos\{(\tau_1 + 2t_2)x + \tau_2 y\} - \cos\{(\tau_1 - 2t_2)x - \tau_2 y\}].$
- (2) $(-1/\tau)(-1)^n \cos \tau x \csc y + \sin(2tx + \tau y) \sec x$
 $+ 2(-1)^{n_1}(-1/\tau_1)[\sin\{(\tau_1 + 2t_2)x + \tau_2 y\} - \sin\{(\tau_1 - 2t_2)x - \tau_2 y\}].$
- (3) $\sin(2tx + \tau y) + (-1)^n(-1/\tau) \cos 2tx \csc y$
 $+ 2(-1)^{n_1}(-1/\tau_1)[\sin\{2t_1 + 2t_2)x + \tau_2 y\} - \sin\{(2t_1 - 2t_2)x - \tau_2 y\}].$
- (4) $(-1/\tau) \cos(2tx + \tau y) \csc x + \sin \tau x \sec y$
 $+ 2(-1/\tau_2)[\sin\{\tau_1 + 2t_2)x + \tau_2 y\} + \sin\{(\tau_1 - 2t_2)x - \tau_2 y\}].$
- (5) $(-1/\tau)[\cos(2tx + \tau y) \sec x + (-1)^n \cos \tau x \sec y]$
 $+ 2(-1)^{n_1}(-1/\tau_1 \tau_2)[\cos\{\tau_1 + 2t_2)x + \tau_2 y\} + \cos\{(\tau_1 - 2t_2)x - \tau_2 y\}].$
- (6) $(-1/\tau)[\cos(2tx + \tau y) + (-1)^n \cos 2tx \sec y]$
 $+ 2(-1)^{n_1}(-1/\tau_1 \tau_2)[\cos\{(2t_1 + 2t_2)x + \tau_2 y\} + \cos\{(2t_1 - 2t_2)x - \tau_2 y\}].$
- (7) $(-1)^n[\sin(2tx + \tau y) \csc x + \sin \tau x \csc y]$
 $- 2(-1)^n[\cos\{(\tau_1 + 2t_2)x + \tau_2 y\} - \cos\{(\tau_1 - 2t_2)x - \tau_2 y\}].$
- (8) $(-1/\tau)[(-1)^n \cos(2tx + \tau y) \sec x + \cos \tau x \sec y]$
 $+ 2(-1)^{n_2}(-1/\tau_1 \tau_2)[\cos\{(\tau_1 + 2t_2)x + \tau_2 y\} + \cos\{(\tau_1 - 2t_2)x - \tau_2 y\}].$

5. There are 4 expansions in which the partition is (III). All are of the form

$$\phi_{abc}(x, y) = T_{abc}(x, y) + 4 \sum_{(m)} q^m (\sum F_{abc}(x, y; m)) + 4 \sum_{(n)} q^{2n} (\sum G_{abc}(x, y; 2n))$$

and as before we need list only 2.

	abc	K_{abc}	$T_{abc}(x, y)$ (Partition III)
(1)	020	$\vartheta'_1 \vartheta_0 \vartheta_2$	$\sec x$
(2)	310	$\vartheta'_1 \vartheta_0 \vartheta_3$	$\csc x$

The corresponding F , G follow.

- (1) $F_{020}(x, y; m) = \sin(\tau x + 2ty) \tan x$
 $+ 2(-1)^{d_1+d_2}[\cos\{(2d_1 + \tau_2)x + 2t_2 y\} - \cos\{(2d_1 - \tau_2)x - 2t_2 y\}];$
 $- G_{020}(x, y; 2n) = (-1)^{d+\delta} \sin 2dx \csc x - \sin(\tau'x + 2t'y) \tan x +$
 $- 2(-1)^{d'+\delta'}[\cos\{(2d'_1 + \tau'_2)x + 2t'_2 y\} - \cos\{(2d'_1 - \tau'_2)x - 2t'_2 y\}].$
- (2) $F_{310}(x, y; m) = (-1/\tau) \cos(\tau x + 2ty) \cot x +$
 $+ 2(-1)^{\delta_1}(-1/\tau_2)[\sin\{(2d_1 + \tau_2)x + 2t_2 y\} + \sin\{(2d_1 - \tau_2)x - 2t_2 y\}];$
 $G_{310}(x, y; 2n) = (-1/\tau') \cos(\tau'x + 2t'y) \cot x + (-1)^{\delta} \sin 2dx \sec x +$
 $+ 2(-1)^{\delta'_1}(-1/\tau'_2)[\sin\{(2d'_1 + \tau'_2)x + 2t'_2 y\} + \sin\{(2d'_1 - \tau'_2)x - 2t'_2 y\}].$

6. For partition (IV) there are 8 expansions, of which only 4 need be tabulated. All are of the form

$$\phi_{abc}(x, y) = 4 \sum_{(m)} q^{m/2} (\sum F_{abc}(x, y; m)).$$

	abc	K_{abc}	(Partition IV)
(1)	202	$\vartheta'_1 \vartheta_2 \vartheta_3.$	
(2)	131	$\vartheta'_1 \vartheta_0 \vartheta_2.$	
(3)	132	$\vartheta'_1 \vartheta_2 \vartheta_3.$	
(4)	201	$\vartheta'_1 \vartheta_0 \vartheta_2.$	

The corresponding F 's follow.

- (1) $\sin tx(\cot x - \tan y)$
 $- 2(-1)^{\delta_2} [\cos\{(t_1 + 2d_2)x + 2\delta_2 y\} - \cos\{(t_1 - 2d_2)x - 2\delta_2 y\}].$
- (2) $(-1/m) \sin tx(\cot x + \cot y)$
 $- 2(-1/m_1) [\cos\{(t_1 + 2d_2)x + 2\delta_2 y\} - \cos\{(t_1 - 2d_2)x - 2\delta_2 y\}].$
- (3) $(-1/t) \cos tx(\tan x + \tan y)$
 $- 2(-1/t_1)(-1)^{a_2 + \delta_2} [\sin\{(t_1 + 2d_2)x + 2\delta_2 y\} - \sin\{(t_1 - 2d_2)x - 2\delta_2 y\}].$
- (4) $(-1/\tau) \cos tx(-\tan x + \cot y)$
 $+ 2(-1/\tau_1)(-1)^{a_2} [\sin\{(t_1 + 2d_2)x + 2\delta_2 y\} - \sin\{(t_1 - 2d_2)x - 2\delta_2 y\}].$

7. For partition (V) there are 4 expansions, of which only 3 need be tabulated. All are of the form

$$\phi_{abc}(x, y) = T_{abc}(x, y) + 4 \sum_{(n)} q^{2n} (\sum F_{abc}(x, y; 2n)),$$

where $T_{abc}(x, y)$ is a trigonometric term independent of the partition.

	abc	K_{abc}	$T_{abc}(x, y)$ (Partition V)
(1)	222	$\vartheta'_1 \vartheta_0 \vartheta_3$	$1 - \tan x \tan y$
(2)	121	$\vartheta'_1 \vartheta_0 \vartheta_3$	$1 + \tan x \cot y$
(3)	211	$\vartheta'_1 \vartheta_0 \vartheta_3$	$-1 + \cot x \cot y.$

The corresponding F 's follow.

- (1) $(-1)^\delta [\sin 2(dx + \delta y) \tan x + (-1)^d \sin 2dx(\tan y - \cot x)] +$
 $+ 2(-1)^{d_1 + \delta_1 + \delta_2} [\cos 2\{(d_1 + d_2)x + \delta_2 y\} - \cos 2\{(d_1 - d_2)x - \delta_2 y\}].$
- (2) $\sin 2(dx + \delta y) \tan x - (-1)^{d+\delta} \sin 2dx(\cot x + \cot y) +$
 $+ 2(-1)^{d_1 + \delta_1} [\cos 2\{(d_1 + d_2)x + \delta_2 y\} - \cos 2\{(d_1 - d_2)x - \delta_2 y\}].$
- (3) $(-1)^\delta \sin 2dx(\cot y - \tan x) + (-1)^d \sin 2(dx + \delta y) \cot x +$
 $- 2(-1)^{d_2 + \delta_2} [\cos 2\{(d_1 + d_2)x + \delta_2 y\} - \cos 2\{(d_1 - d_2)x - \delta_2 y\}].$

8. It will be sufficient to indicate how one of the foregoing expansions can be calculated. We have, for example,

$$\phi_{223}(x, y) = \vartheta'_1 \frac{\vartheta_2(x+y)}{\vartheta_0(x)\vartheta_3(y)} \cdot \vartheta_0\vartheta_2 \frac{\vartheta_0(x)}{\vartheta_2(x)} = \vartheta'_1\vartheta_0\vartheta_2\psi_{223}(x, y).$$

The expansions of the two functions which are multiplied are given in reduced form in the papers cited in § 1.† Note, however, that we also have,

$$\phi_{223}(x, y) = \vartheta'_1 \frac{\vartheta_2(x+y)}{\vartheta_2(x)\vartheta_1(y)} \cdot \vartheta_0\vartheta_2 \frac{\vartheta_1(y)}{\vartheta_3(y)} = \vartheta'_1\vartheta_0\vartheta_2\psi_{223}(x, y);$$

this gives an expansion corresponding to partition (IV) whereas the preceding identity gives one belonging to partition (I). We are thus led to *arithmetically* different expansions for the same function, which is, of course, desirable from the point of view of the applications. Similarly all 48 expansions can be calculated. In what follows we list the expansions obtained by the second method of calculation; it will be sufficient to give the value of $F_{abc}(x, y; n)$ since K_{abc} and $T_{abc}(x, y)$ are the same as before.

9. The expansions corresponding to partition (I) with

$$\phi_{abc}(x, y) = 4 \sum_{(m)} q^{m/2} (\sum F_{abc}(x, y; m))$$

are as follows:

abc	$F_{abc}(x, y; m)$
130	$(-1/m)\sin(tx + \tau y)$ $+ 2(-1/\tau_1)(-1/m_2)[\sin\{t_2x + (\tau_2 + 2t_1)y\} + \sin\{t_2x + (\tau_2 - 2t_1)y\}].$
233	$(-1/\tau)\cos(\tau x + ty)$ $+ 2(-1)^{n_1}(-1/\tau_1\tau_2)[\cos\{\tau_2x + (t_2 + 2t_1)y\} + \cos\{\tau_2x + (t_2 - 2t_1)y\}].$
000	$\sin ty \csc y - 2[\cos\{2t_1x + (\tau_1 + t_2)y\} - \cos\{2t_1x + (\tau_1 - t_2)y\}].$
030	$(-1/\tau)\cos ty \sec y$ $+ 2(-1)^{n_1}(-1/\tau_1\tau_2)[\cos\{2t_1x + (\tau_1 + t_2)y\} + \cos\{2t_1x + (\tau_1 - t_2)y\}].$
303	$(-1/t)\cos ty \sec y$ $+ 2(-1/\tau_1t_2)[\cos\{2t_1x + (\tau_1 + t_2)y\} + \cos\{2t_1x + (\tau_1 - t_2)y\}].$
333	$(-1/m)\sin ty \csc y$ $- 2(-1)^{n_1}(-1/m_2)[\cos\{2t_1x + (\tau_1 + t_2)y\} - \cos\{2t_1x + (\tau_1 - t_2)y\}].$
202	$(-1/\tau)\cos(tx + \tau y)\sec y$ $+ 2(-1/\tau_1\tau_2)[\cos\{t_2x + (\tau_2 + \tau_1)y\} + \cos\{t_2x + (\tau_2 - \tau_1)y\}].$

† See also, E. T. Bell, *Giornale di Matematiche*, vol. 61 (1921), pp. 93-114.

- 131 $(-1/m) \sin(tx + \tau y) \csc y$
 $- 2(-1)^{n_1}(-1/m_2) [\cos\{t_2x + (\tau_2 + \tau_1)y\} - \cos\{t_2x + (\tau_2 - \tau_1)y\}].$
- 300 $(-1/\tau) \cos ty \sec y$
 $+ 2(-1/\tau_1\tau_2) [\cos\{2t_1x + (\tau_1 + t_2)y\} + \cos\{2t_1x + (\tau_1 - t_2)y\}].$
- 033 $(-1/t) \cos ty \sec y$
 $+ 2(-1)^{n_1}(-1/\tau_1t_2) [\cos\{2t_1x + (\tau_1 + t_2)y\} + \cos\{2t_1x + (\tau_1 - t_2)y\}].$
- 132 $(-1/m) \sin(tx + \tau y) \sec y$
 $+ 2(-1/\tau_1m_2) [\sin\{t_2x + (\tau_2 + \tau_1)y\} + \sin\{t_2x + (\tau_2 - \tau_1)y\}].$
- 201 $(-1/\tau) \cos(tx + \tau y) \csc y$
 $+ 2(-1)^{n_1}(-1/\tau_2) [\sin\{t_2x + (\tau_2 + \tau_1)y\} - \sin\{t_2x + (\tau_2 - \tau_1)y\}].$
- 203 $(-1/\tau) \cos(tx + \tau y)$
 $+ 2(-1/\tau_1\tau_2) [\cos\{t_2x + (\tau_2 + 2t_1)y\} + \cos\{t_2x + (\tau_2 - 2t_1)y\}].$

10. The expansions in this section correspond to partition (II) with

$$\phi_{abc}(x, y) = T_{abc}(x, y) + 4 \sum_{(n)} q^n (\sum F_{abc}(x, y; n)).$$

abc

$F_{abc}(x, y; n)$

- 011 $\sin(\tau x + 2ty) \csc y + \sin \tau y \csc x$
 $- 2[\cos\{\tau_2x + (2t_2 + \tau_1)y\} - \cos\{\tau_2x + (2t_2 - \tau_1)y\}].$
- 021 $(-1)^n [(-1/\tau) \cos(\tau x + 2ty) \csc y + \sin \tau y \sec x]$
 $+ 2(-1)^n (-1/\tau_2) [\sin\{\tau_2x + (2t_2 + \tau_1)y\} - \sin\{\tau_2x + (2t_2 - \tau_1)y\}].$
- 312 $(-1)^n \sin(\tau x + 2ty) \sec y + (-1/\tau) \cos \tau y \csc x$
 $+ 2(-1)^{n_2} (-1/\tau_1) [\sin\{\tau_2x + (2t_2 + \tau_1)y\} + \sin\{\tau_2x + (2t_2 - \tau_1)y\}].$
- 322 $(-1/\tau) [\cos(\tau x + 2ty) \sec y + (-1)^n \cos \tau y \sec x]$
 $+ 2(-1)^{n_1} (-1/\tau_1\tau_2) [\cos\{\tau_2x + (2t_2 + \tau_1)y\} + \cos\{\tau_2x + (2t_2 - \tau_1)y\}].$
- 020 $(-1/\tau) [(-1)^n \cos(\tau x + 2ty) + \cos 2ty \sec x]$
 $+ 2(-1)^{n_2} (-1/\tau_1\tau_2) [\cos\{\tau_2x + (2t_2 + \tau_1)y\} + \cos\{\tau_2x + (2t_2 - \tau_1)y\}].$
- 310 $(-1)^n \sin(\tau x + 2ty) + (-1/\tau) \cos ty \csc x$
 $+ 2(-1)^{n_1} (-1/\tau_2) [\sin\{\tau_1x + (2t_1 + 2t_2)y\} + \sin\{\tau_1x + (2t_1 - 2t_2)y\}].$
- 311 $(-1)^n [\sin(\tau x + 2ty) \csc y + \sin \tau y \csc x]$
 $- 2(-1)^n [\cos\{\tau_2x + (2t_2 + \tau_1)y\} - \cos\{\tau_2x + (2t_2 - \tau_1)y\}].$
- 022 $(-1/\tau) [(-1)^n \cos(\tau x + 2ty) \sec y + \cos \tau y \sec x]$
 $+ 2(-1)^{n_2} (-1/\tau_1\tau_2) [\cos\{\tau_2x + (2t_2 + \tau_1)y\} + \cos\{\tau_2x + (2t_2 - \tau_1)y\}].$

11. The expansions given below correspond to partition (III) with

$$\phi_{abc}(x, y) = T_{abc}(x, y) + 4 \sum_{(m)} q^m (\sum F_{abc}(x, y; m))$$

$$+ 4 \sum_{(n)} q^{2n} (\sum G_{abc}(x, y; 2n)).$$

We have for $(abc) = (031)$ or (332) :

$$\begin{aligned}
 -F_{031}(x, y; m) &= (-1/\tau) \cos(2tx + \tau y) \cot y \\
 &\quad + 2(-1)^{\delta_1} (-1/\tau) [\sin\{2t_2x + (\tau_2 + 2d_1)y\} - \sin\{2t_2x + (\tau_2 - 2d_1)y\}]. \\
 G_{031}(x, y; 2n) &= (-1)^{\delta} \sin 2dy \sec y + (-1/\tau') \cos(2t'x + \tau'y) \cot y + \\
 &\quad + 2(-1)^{\delta_1} (-1/\tau'_2) [\sin\{2t'_2x + (2d'_1 + \tau'_2)y - \sin\{2t'_2x + (\tau'_2 - 2d'_1)y\}]. \\
 -F_{332}(x, y; m) &= \sin(2tx + \tau y) \tan y \\
 &\quad + (-1)^{\delta_1 + \delta_1} [\cos\{2t_2x + (\tau_2 + 2d_1)y\} - \cos\{2t_2x + (\tau_2 - 2d_1)y\}]. \\
 G_{332}(x, y; 2n) &= -(-1)^{\delta + \delta} \sin 2dy \csc y + \sin(2t'x + \tau'y) \tan y + \\
 &\quad + 2(-1)^{\delta_1 + \delta_1} [\cos\{2t'_2x + (\tau'_2 + 2d'_1)y\} - \cos\{2t'_2x + (\tau'_2 - 2d'_1)y\}].
 \end{aligned}$$

12. The following correspond to partition (IV) with

$$\phi_{abc}(x, y) = 4 \sum_{(m)} q^{m/2} (\sum F_{abc}(x, y; m)).$$

\overline{abc}	$\overline{F_{abc}(x, y; m)}$
110	$(\cot x + \cot y) \sin ty$ $- 2[\cos\{2d_2x + (2\delta_2 + t_1)y\} - \cos\{2d_2x + (2\delta_2 - t_1)y\}].$
120	$(-1/\tau) (\tan x + \tan y) \cos ty$ $- 2(-1/\tau_1) (-1)^{\delta_2 + \delta_2} [\sin\{2d_2x + (2\delta_2 + t_1)y\} + \sin\{2d_2x + (2\delta_2 - t_1)y\}].$
213	$(-1/t) (\cot x - \tan y) \cos ty$ $+ 2(-1/t_1) (-1)^{\delta_2} [\sin\{2d_2x + (2\delta_2 + t_1)y\} + \sin\{2d_2x + (2\delta_2 - t_1)y\}].$
223	$(-1/m) (-\tan x + \cot y) \sin ty$ $- 2(-1/m_1) (-1)^{\delta_2} [\cos\{2d_2x + (2\delta_2 + t_1)y\} - \cos\{2d_2x + (2\delta_2 - t_1)y\}].$

13. Finally, the expansions corresponding to partition (V) with

$$\phi_{abc}(x, y) = T_{abc}(x, y) + 4 \sum_{(n)} q^{2n} (\sum F_{abc}(x, y; 2n)),$$

are:

\overline{abc}	$\overline{F_{abc}(x, y; 2n)}$
(222)	$(-1)^{\delta} [\sin 2(dx + \delta y) \tan y + (-1)^{\delta} \sin 2dy (\tan x - \cot y)] +$ $+ 2(-1)^{\delta_1 + \delta_2 + \delta_2} [\cos 2\{d_1x + (\delta_1 + d_2)y\} - \cos 2\{d_1x + (\delta_1 - d_2)y\}].$
(121)	$(-1)^{\delta} [-(-1)^{\delta} \sin 2(dx + \delta y) \cot y + \sin 2dy (\tan x + \tan y)] +$ $+ 2(-1)^{\delta_1 + \delta_1 + \delta_2} [\cos 2\{d_1x + (\delta_1 + d_2)y\} - \cos 2\{d_1x + (\delta_1 - d_2)y\}].$
(211)	$(-1)^{\delta} [\sin 2(dx + \delta y) \cot y + \sin 2dy (\cot x - \tan y)] +$ $- 2(-1)^{\delta_1 + \delta_2} [\cos 2\{d_1x + (\delta_1 + d_2)y\} - \cos 2\{d_1x + (\delta_1 - d_2)y\}].$

CONCERNING CONTINUA OF FINITE DEGREE AND LOCAL SEPARATING POINTS.

By G. T. WHYBURN.

A point p of a compact metric continuum M is said to be a point of finite degree * of M provided that for each $\epsilon > 0$ there exists an uncountable family G of neighborhoods of p in M each of diameter $< \epsilon$ and each having a finite set for a boundary and such that for any two neighborhoods U, V of G we have either $\bar{U} \subset V$ or $\bar{V} \subset U$. Similarly, p is said to be of degree α provided α is the least cardinal number such that for each ϵ the neighborhoods of G can be chosen so that the boundary of each of them is of power $\leq \alpha$. (We shall be concerned principally with the case where α is an integer.) If X is a given subset of M , then p will be said to be of finite degree or of degree α in M relative to X provided that, in addition to satisfying the conditions previously stated, the neighborhoods of G can be so chosen that the boundary of each of them is a subset of X . These notions concerning the degree of a point should not be confused with the analogous but distinctly different notions concerning the order or index of a point as introduced by Menger and Urysohn. Both concepts will be used in this paper, but careful reading should enable one to avoid confusing them.

A continuum is said to be of finite degree provided each of its points is a point of finite degree. In this paper we shall show, among other results, that such continua may be identified with the continua previously studied by the author † which have the property that any subcontinuum contains uncountably many local separating points ‡ of the given continuum.

(1) THEOREM. *All save possibly a countable number of the local separating points of any continuum M are points of degree 2 of M .* §

* See H. Kamiya, *Tôhoku Mathematical Journal*, vol. 36 (1933), pp. 58-72.

† See my paper "Decompositions of continua by means of local separating points," *American Journal of Mathematics*, vol. 55 (1933), pp. 437-457.

‡ The point p of a continuum M is called a local separating point of M provided there exists a neighborhood R of p in M such that p separates \bar{R} between some pair of points of the component of \bar{R} containing p . See my paper "Local separating points of continua," *Monatshefte für Mathematik und Physik*, vol. 36 (1929), pp. 305-314.

§ This theorem is closely related to Theorem 9 of my paper "Local separating points of continua," *loc. cit.*; compare also Kamiya, *loc. cit.*, Theorem 18.

Suppose on the contrary, that there exists an uncountable set H of local separating points of M no one of which is a point of degree 2. Then applying Theorem 10 of my paper just referred to, we obtain a connected open subset R of M bounded by two points α and β of H and such that α and β are separated in \bar{R} by each point of an uncountable subset E of H . Let E_0 be the set of all points of E which are condensation points of E . Then since each point of E is a cut point of \bar{R} it follows by a theorem of the author's* that E_0 contains a subset S such that $E_0 - S$ is countable and each point of S is a point of order 2 in \bar{R} relative to S .

Now let p be any point of S and let ϵ be any positive number. There exists an ϵ -neighborhood U of p in M bounded by two points x and y of S . Let K be the set of all points in U which separate x and y in \bar{U} . Then $K \supset S \cdot U$ and the set $K + x + y$ can be † ordered in a natural linear order from x to y . Then since x and y are points of order 2 relative to S and every point of S is a condensation point of S and since furthermore K is the sum of a G_δ -set and a countable set,‡ it follows that there exist 0-dimensional perfect sets A and B in K such that if $a \in A$ and $b \in B$, we have the order $x < a < p < b < y$. Clearly there exists a $(1-1)$ order reversing correspondence (a, b) between A and B . Then each pair of corresponding points (a, b) determines an ϵ -neighborhood $U(a, b)$ in M bounded by $a + b$; and if (a_1, b_1) and (a_2, b_2) are any two such pairs and $a_1 < a_2$, we have $U(a_1, b_1) \supset U(a_2, b_2)$. Thus p is of degree 2 in M , contrary to supposition; and our theorem is proven.

(1.1) *Any point of finite degree or of degree n is of the same degree in M relative to the local separating points of M of degree 2.*

For if p is of finite degree (or of degree n) in M , then for each $\epsilon > 0$ we have an uncountable family G of ϵ -neighborhoods satisfying the conditions of the definition and such that the boundary of each neighborhood is finite (or of power $\leq n$). Now if we choose the neighborhoods U of G so that every point of the boundary of U is a limit point of $M - U$, as we clearly can do, then every boundary point of a neighborhood of G is a local separating point of M . Then since by (1) all save a countable number of the local separating points of M are points of degree 2, it follows that for all save a

* See my paper in *Transactions of the American Mathematical Society*, vol. 33 (1931), p. 450.

† See my paper, *loc. cit.*, p. 446.

‡ See my paper in *Transactions of the American Mathematical Society*, vol. 32 (1930), p. 151.

countable number of the neighborhoods U of G , $F(U)$ consists wholly of local separating points of M of degree 2; and our result follows at once.

Throughout the rest of the paper, if M is a continuum then we shall denote by L the set of all local separating points of M and by Q the set of all points of L which are of degree 2 in M . Now from (1) and (1.1) we have at once

(1.2) *Every point Q is of degree 2 in M relative to Q and $L - Q$ is countable.*

Now since every countable set which separates a continuum M contains at least one local separating point of M ,* a slight modification of the argument given for (1.1) suffices to prove

(1.3) *Every point of countable degree (i. e., $\leq \aleph_0$) of a continuum M is a condensation point of local separating points of degree 2 of M .*

(2) THEOREM. *If p is a point of finite order or of order n in a compact continuum M relative to Q , then p is of the same degree in M relative to Q .*

Proof. Given $\epsilon > 0$, there exists an $\epsilon/3$ -neighborhood R of p whose boundary consists of a finite number, k , of points p_1, p_2, \dots, p_k of Q where $k = n$ in case p is of order n relative to Q . For each $i \leq k$ there exists an $\epsilon/3$ -neighborhood V_i of p_i whose boundary consists of just two points x_i and y_i one of which, x_i , lies in R and the other in $M - \bar{R}$, and these can be chosen so that $\bar{V}_i \cdot \bar{V}_j = 0$, $i \neq j$, and $\bar{V}_i \cdot p = 0$. Now since for each $i \leq k$, p_i is of degree 2, it follows that if K_i denotes the set of all points separating y_i and p_i in \bar{V}_i , then K_i is uncountable and, in fact, contains a 0-dimensional perfect set A_i [see the proof of (1)]. The sets K_i are ordered from p_i to y_i as in (1); and clearly there exists a $(1-1-1-\dots-1)$ order preserving correspondence $(a_1, a_2, a_3, \dots, a_k)$ between the sets A_1, A_2, \dots, A_k , i. e., a grouping of the points of these sets into groups (a_1, a_2, \dots, a_k) , $a_i \in A_i$ such that when $a_1^1, a_1^2 \in A_1$ and $a_1^1 < a_1^2$ in K_1 , then for each $i \leq k$, $a_i^1 < a_i^2$ in K_i . It is seen at once that each such group (a_1, a_2, \dots, a_k) is the boundary of an ϵ -neighborhood $U(a_1, \dots, a_k)$ of p obtained by adding to R certain points of the neighborhoods V_i and that $a_1^1 < a_1^2$, $(a_1^1, a_1^2 \in A_1)$, implies $\overline{U(a_1^1, a_2^1, \dots, a_k^1)} \subset \overline{U(a_1^2, a_2^2, \dots, a_k^2)}$. Thus p is of the same degree in M as its order in M relative to Q ; and hence, by (1.1), it is likewise of the same degree in M relative to Q .

* See my paper "Local separating points of continua," *loc. cit.*

(2.1) COROLLARY. *The points of finite degree or of degree n of a continuum M are exactly the points of finite order or of order n in M relative to Q .*

(2.2) (Kamiya, *loc. cit.*) *The points of M of finite degree or of degree $\leq n$ in M form a G_δ -set.*

(2.3) *Any point of a compact continuum M of infinite degree lies in a non-degenerate continuum of such points.**

This is an immediate consequence of (2.1) and of a theorem of Menger's.†

(2.4) *In order that a compact continuum M be of finite degree it is necessary and sufficient that any two points of M be separated in M by a finite number of points of Q .*

This can be proven at once from (2) by a simple application of the Borel Theorem.

(3) THEOREM. *In order that a point p be a point of finite degree of a compact continuum M it is necessary and sufficient that $\dim_p(M - Q + p) = 0$.*

The necessity of the condition follows at once from (1.2). To prove the sufficiency, let us suppose $\dim_p(M - Q + p) = 0$ and $\epsilon > 0$. Then there exists an $\epsilon/3$ -neighborhood R of p with $F(R) \subset Q$. Now for each $x \in F(R)$, it follows by (1.3) that there exists an $\epsilon/3$ -neighborhood U_x of x such that $F(U_x)$ consists of exactly two points of Q . Applying the Borel Theorem we obtain a finite number of the neighborhoods U_x , say U_1, U_2, \dots, U_n whose sum covers $F(R)$. Then if $U = R + U_1 + U_2 + \dots + U_n$, it is clear that U is an ϵ -neighborhood of p whose boundary consists of n or less points of Q . Thus p is of finite order in M relative to Q , and hence by (2) it is also of finite degree in M .

(3.1) COROLLARY. *In order that a continuum M be of finite degree it is necessary and sufficient that $\dim(M - Q) = 0$.*

Note. It follows at once from this that if M^2 denotes the set of all points of M of order 2, then $\dim(M - M^2) = 0$, because $M - M^2 \subset M - Q$. (See Kamiya, *loc. cit.*) However, examples are easily constructed to show that the condition that $\dim(M - M^2) = 0$ is not sufficient to make M of finite degree.

(4) THEOREM. *In order that a compact continuum M be of finite degree*

* Compare with Kamiya, *loc. cit.*, Theorem 10.

† See *Kurventheorie*, Teubner, 1928, p. 128.

at a point p it is necessary and sufficient that every subcontinuum of M containing p contain uncountably many local separating points of M .

The condition is sufficient. For suppose the condition is satisfied at p . Then if p were not a point of finite degree, by (2.2) there would exist a non-degenerate subcontinuum N of M containing p no point of which is of finite degree. But this is impossible, since by hypothesis N must contain uncountably many local separating points of M ; and hence, by (1), it contains at least one point of degree 2.

The condition is also necessary. For let p be of finite degree and let N be any subcontinuum of M containing p . Then there exists an uncountable family $[U]$ of neighborhoods of p of the type we have been considering such that for any $U \in [U]$, $\delta(U) < \delta(N)$ and $F(U)$ is a finite set of local separating points of M . But clearly for each U we have $F(U) \cdot N \neq 0$, and hence N contains uncountably many local separating points of M .

(4.1) COROLLARY. *In order that a compact continuum M be of finite degree it is necessary and sufficient that every subcontinuum of M contain uncountably many local separating points of M .*

Now the property involved in this corollary is readily seen to be cyclicly extensible and reducible.* Hence we have

(4.2) *The property of being a continuum of finite degree is cyclicly extensible and reducible.*

Following the notation used in my paper "Decompositions of continua by means of local separating points" (*loc. cit.*), for any point p of a continuum M let us denote by $C_1(p)$ the maximal subcontinuum of M containing p and containing only a countable number of local separating points of M . Then by (4) we have

(4.3) *In order that p be a point of finite degree of a compact continuum M it is necessary and sufficient that $C_1(p) = p$.*

Now it was shown in the author's paper just mentioned that the decomposition of M into the sets $[C_1(p)]$ is upper semi-continuous† and that the hyperspace C_1 of this decomposition has the property that any sub-

* See Kuratowski and Whyburn, *Fundamenta Mathematicae*, vol. 16 (1930), pp. 305-331.

† See R. L. Moore, "Foundations of point set theory," *American Mathematical Society Colloquium Publications*, 1932, Chapter V.

continuum of C_1 contains uncountably many local separating points of C_1 . Thus by (4.1), C_1 is of finite degree. Thus we have

(4.4) *Every compact continuum M admits of an upper semi-continuous decomposition into continua with a continuum of finite degree as hyperspace.*

(The hyperspace reduces to a single point if and only if the local separating points of M are countable).

(5) THEOREM. *If M is a regular curve of order ≤ 3 , then degree and order are identical for all points of M .*

For let p be a point of such a curve M , let k be the order of p and let $\epsilon > 0$. Then there exists an $\epsilon/3$ -neighborhood R of p whose boundary consists of k points p_1, \dots, p_k , ($k \leq 3$), each point of which is a limit point of $M - R$. For each $i \leq k$, p_i is a point of order 1 either of \bar{R} or of $\overline{M - R}$, say of \bar{S} . Then there exists an $\epsilon/3$ -neighborhood V_i of p_i whose boundary contains just one point q_i of \bar{S} and such that $\bar{V}_i \cdot \bar{V}_j = 0$, $i \neq j$ and $\bar{V}_i \cdot p = 0$. Now since no two true cyclic elements of \bar{V}_i can have a common point* and since p_i is a point of order 1 of \bar{S} , it follows at once that p_i and q_i are separated in \bar{V}_i by uncountably many distinct points. Since all such points are local separating points of M , it results from (1) that p_i and q_i are separated in \bar{V}_i by at least one point x_i of degree 2 (i. e., a point of the set Q). It is easily seen then that the points x_1, \dots, x_k bound an ϵ -neighborhood U of p . Thus p is of order k in M relative to Q and accordingly, by (2), p is of degree k .

In contrast to the theorem just proven, it is notable that *there exist regular curves of order 4 which are of degree \mathfrak{c} at every point*, where \mathfrak{c} denotes the power of the continuum. For the Sierpinski triangle curve † is an example of such a curve. To see that every point of this curve is of degree \mathfrak{c} , in view of (1.3) we have only to note that its local separating points are countable.

THE UNIVERSITY OF VIRGINIA.

* See my paper in the *Bulletin of the American Mathematical Society*, vol. 35 (1929), p. 223.

† See *Comptes Rendus*, vol. 162, p. 629.

THE TOPOLOGY OF (PATH) SURFACES.

By CHARLES B. MORREY, JR.†

In the author's recent work on the analytic characterization of surfaces of finite area, a considerable number of difficulties were encountered, due to the complicated topological structure of the most general surface. The object of the present paper is to study this structure and to provide a systematic treatment of a number of concepts and results which facilitate the discussion of the general surface.

Before giving a definition of what we mean by a (path) surface, we must introduce the notation. We shall use the ordinary notations for the sum and product of point sets. If S is any set, we shall designate by \bar{S} the set S plus its limit points and by S^* the frontier points of S . If P is a point and S a set, we write $P \in S$ instead of " P belongs to S " and $P \notin S$ instead of " P does not belong to S ." We write $\Sigma \subset S$ instead of Σ is a subset of S . By a Jordan region r (in n -space) we mean the interior of a simple closed $(n-1)$ -manifold; we shall frequently also refer to \bar{r} as a Jordan region. We designate the interior of a circle with center at P and radius r by $C(P, r)$.

Further, we shall use a vector notation throughout: the letters u and U will stand for the coördinate vector of a point in a set on which a surface is represented and the letters x and X for that of a point in the (N dimensional) space in which the surface lies. $U(u)$, $x(u)$, $X(U)$, etc. will be vector functions in these spaces; $U_1 \pm U_2$, $x_1 \pm x_2$, etc., will denote the sum or difference vectors, and $|\phi|$ the length of the vector ϕ . If P is in the U space, for instance, U_P will denote the U -coördinate vector of P . The distance between two points P and Q will be denoted by $d(P, Q)$ or by $|u_P - u_Q|$ or $|U_P - U_Q|$ if P and Q are in the u or U space respectively. *All vector functions will be assumed to be continuous.*

Now, let $x_1(u)$ and $X_2(U)$, $u \in \bar{r}$, $U \in \bar{R}$ (Jordan regions in n -space), be two vector functions. Let T , $T: U = U(u)$, be a 1-1 continuous (sense preserving if desired) transformation of \bar{r} into \bar{R} . Define $x_2(u) = X_2[U(u)]$ and $D_T(x_1, X_2)$ as the maximum of $|x_1(u) - x_2(u)|$ for $u \in \bar{r}$. Then we define $\|x_1, X_2\|$ as the greatest lower bound of $D_T(x_1, X_2)$ for all T . By a (path) manifold, M , we shall mean a class of equations $x = x(u)$, such that if $x_1(u)$, and $X_2(U)$ are any two of the $x(u)$, then $\|x_1, X_2\| = 0$. Any one of the equations of the class will be called a representation of the manifold.

† National Research Fellow (1931-1933).

It is easy to see that if $x_1(u)$, $X_2(U)$, and $\hat{X}_3(\hat{U})$, $u \in \bar{r}$, $U \in \bar{R}$, $\hat{U} \in \bar{\hat{R}}$, then $\|x_1, \hat{X}_3\| \leq \|x_1, X_2\| + \|X_2, \hat{X}_3\|$, and also that $\|x_1, X_2\| = \|X_2, x_1\|$. Thus if $\|x_1, X_2\| = \|X_2, \hat{X}_3\| = 0$, then $\|x_1, \hat{X}_3\| = 0$. If $n = 1$, the manifold will be called a *curve*, and if $n = 2$, it will be called a *surface*. It is clear that a manifold is defined by any one of its representations. This definition of manifold is clearly the logical consequence of Fréchet's † definition of the *distance between two surfaces* (or manifolds). This may be defined as follows: Let S_1 be defined by the equation $x = x_1(u)$, $u \in \bar{r}$, and S_2 by $x = X_2(U)$, $U \in \bar{R}$; then the Fréchet distance $\|S_1, S_2\| = \|x_1, X_2\|$, it being clear from the above that this is independent of the particular representations. Clearly $\|S_1, S_2\| = 0$ when and only when S_1 and S_2 are identical.

I shall now merely give information enabling one to define a point of a surface. Let S , $S: x = x(u)$, $u \in \bar{r}$, be a surface and let $u_1 \neq u_2$ in \bar{r} . Then the point P_1 of S defined by $x = x(u_1)$ will coincide with P_2 , that defined by $x = x(u_2)$, if and only if $x(u)$ is constant over a continuum of \bar{r} containing u_1 and u_2 . The fact that (see § 2) the collection of maximal continua over each of which $x(u)$ is constant is an "upper semicontinuous collection (see § 2) of mutually exclusive continua filling up \bar{r} ," suggests an intimate relation between the structure of a surface and the theory of these collections developed by R. L. Moore. With his definition of limit element in such a collection, it is seen that there is a 1 — 1 continuous correspondence between the continua of the above collection and the points of S . Furthermore, it will appear later that if $x = x(u)$, $u \in \bar{r}$, and $x = X(U)$, $U \in \bar{R}$, are representations of the same surface, then there is a 1 — 1 continuous correspondence between the collections of continua in \bar{r} and \bar{R} , each one of which corresponds to a point of S , so that the definition of identity or distinctness of two points of S is independent of the particular representation.

By a non-degenerate surface, we shall mean one which possesses a representation $x = x(u)$, $u \in \bar{r}$, in which $x(u)$ is not constant over any continuum of \bar{r} (i. e. the points of S are in a 1 — 1 continuous correspondence with those of \bar{r}). Unfortunately not all surfaces are non-degenerate. To study the structure of the general surface, the notion of a "hemicactoid" is introduced. It is shown that if \bar{H} is a hemicactoid and $X(U)$ is a (continuous) vector function defined on it, then (1) there exists a "monotone" transformation (carrying only continua into points) $U = U(u)$, $u \in \bar{r}$, of \bar{r} into \bar{H} (which transformation also satisfies another interesting condition) and (2) if we define $x(u) = X[U(u)]$, then all the surfaces $x = x(u)$ thus obtained are

† M. Fréchet, "Sur la distance de deux surfaces," *Annales de la Société Polonaise de Mathématiques*, vol. 3 (1924), pp. 4-19.

identical. We then introduce parametric representations of surfaces on these hemicactoids and it is shown that every surface can be represented on some hemicactoid \bar{H} in such a way that the representing vector function is not constant over any continuum of \bar{H} . Some useful theorems about the relations between two different representations of the same surface are demonstrated. Finally, a criterion, in terms of its given representation, that a surface be non-degenerate is developed and it is shown that a surface "bounded" by a Jordan curve and represented by a vector function, all of whose components are monotone in the sense of Lebesgue, is non-degenerate.

1. *On continuous curves in general.* In this section we shall recall a number of general definitions and results about continuous curves and shall develop a representation of a continuous curve which will be of use in later work. This representation is entirely analogous to a representation of an acyclic (see Definition 14, below) continuous curve developed by R. L. Wilder,[†] the arcs of his representation being replaced, in the present one, by simple cyclic chains (see below) of the continuous curve. All point sets considered in this section will be supposed to be in a bounded portion of Euclidean n -space.

Definitions 1-16 and Lemmas 1-7 are all either to be found in the literature or are easily deducible from known results. They are merely included to furnish an adequate introduction to the terminology and certain of the results of modern point set theory.

Definition 1. By a *maximal* set possessing a given property, we shall mean a set which is not a proper subset of any other set possessing the given property.

Definition 2. By a *component* of a point set, S , we shall mean a maximal connected subset of S (i. e. a maximal set possessing the property of being a connected subset of S).

Definition 3. A set is said to be *completely disconnected* if all of its components are points.

Definition 4. A point P of a connected set S is said to be a *cut point* of S if $S - P$ is not connected. Two sets H and K are said to be *mutually separated* if they have no points in common and neither contains a limit point of the other. If H, K , and T are proper subsets of the connected point set M , then T is said to *separate* H from K in M if $M - T$ is the sum of two mutually

[†] R. L. Wilder, "Concerning continuous curves," *Fundamenta Mathematicae*, vol. 7 (1925), p. 365.

separated point sets containing H and K respectively. In particular, H , K , and T may be points.

Definition 5.[†] A point, P , of a continuum \bar{C} , is said to be an *end point* of \bar{C} provided that if \bar{c} is a subcontinuum of \bar{C} containing P , then P is not a limit point of any connected subset of $\bar{C} - \bar{c}$.

Definition 6. A point set D , is said to be *open with respect to a set S* , if $D = S \cdot G$, where G is an open subset of the given n -space.

Definition 7.[§] The point set S , is said to be *connected in kleinen* at the point P , if $P \in S$ and every subset D , of S , which is open in S and contains P contains a subset d , having all these properties and which is a subset of a component of D .

Definition 8.[§] The point set S , is said to be *locally connected* at the point P , if $P \in S$ and every subset D , of S , which is open in S and contains P contains a connected subset d , with all of these properties.

Definition 9.[§] A *continuous curve* is a continuum which is connected in kleinen at each of its points.

LEMMA 1. Any two points of a connected subset D , of a continuous curve \bar{M} , which is open in \bar{M} may be joined by a simple arc which is a subset of D .

Proof. This is a subcase of Theorem 10, chapter II, P. S. T.

LEMMA 2. A necessary and sufficient condition that a bounded continuum \bar{M} , be a continuous curve is that it should be uniformly arcwise connected in kleinen, i. e. for every $\epsilon > 0$, there exists a $\delta > 0$ such that if P and Q are any two points of \bar{M} at a distance $< \delta$ apart, they can be joined by an arc of \bar{M} of diameter $< \epsilon$.

Proof. This theorem is well known. The necessity of the condition follows easily from the above lemma and the sufficiency from the Definitions 7, 8, and 9.

LEMMA 3. A necessary and sufficient condition that a bounded continuum \bar{M} be a continuous curve is that there exists a continuous vector

[†] G. T. Whyburn, "Concerning the structure of a continuous curve," *American Journal of Mathematics*, vol. 50 (1928), pp. 167-194, § 1. This paper will hereafter be referred to as W.

[§] G. T. Whyburn, "Concerning continua in the plane," *Transactions of the American Mathematical Society*, vol. 29 (1929) p. 382.

§ R. L. Moore, "Foundations of point set theory," *Colloquium Publications*, vol. 13, p. 94. We shall hereafter refer to this book as P. S. T.

function $x(u)$, $u \in \bar{r}$, \bar{r} being a Jordan region in k -space, such that the transformation $x = x(u)$, carries \bar{r} into \bar{M} .

Proof. This theorem was originally proved for the case that \bar{r} is an interval.† It is easy to see that any k dimensional Jordan region is a continuous curve, and also that there exists a continuous transformation of such a region into an interval. Hence the above theorem follows.

Definition 10 (P. S. T., p. 63 and p. 72). If \bar{M} is a continuum and P is a point of \bar{M} such that there do not exist two points A and B such that (1) P separates A and B in \bar{M} and (2) P is the only point separating A and B in \bar{M} , then P is a proper point of \bar{M} and the set of all points X of \bar{M} which are not separated from P in \bar{M} by a point of \bar{M} is called a *simple link* of \bar{M} . A simple link containing two points will be called *non-degenerate*.

Definition 11 (W., § 1). A continuous curve \bar{M} , is said to be *cyclically connected* if every two of its points lie on a simple closed curve of \bar{M} . A *maximal cyclic curve* of \bar{M} is a maximal set \bar{C} , possessing the properties: (1) \bar{C} is a cyclically connected continuous curve and (2) \bar{C} is a subset of \bar{M} .

Definition 12 (W., § 1). By an *internal point* of a maximal cyclic curve \bar{C} , of \bar{M} is meant a point of \bar{C} which is not a cut point of \bar{M} .

Definition 13 (W., § 1). A subset E , of \bar{M} is called a *cyclic element* of \bar{M} if it belongs to one of the following three classes: (a) maximal cyclic curves of \bar{M} , (b) cut points of \bar{M} , and (c) end points of \bar{M} . Those of class (a) constitute the *non-degenerate cyclic elements* of \bar{M} .

LEMMA 4 (P. S. T., chapter II, Theorems 68-70). *Every simple link of a continuous curve \bar{M} , is a cyclic element of \bar{M} . Every non-degenerate cyclic element of \bar{M} , is a simple link of \bar{M} . A necessary and sufficient condition that a degenerate cyclic element P , of \bar{M} be a simple link of \bar{M} is that it does not belong to any non-degenerate simple link of \bar{M} .*

LEMMA 5.† *Every continuous curve is the sum of its cyclic elements.*

LEMMA 6 (W., § 1). (1) *No two cyclic elements of \bar{M} have more than one point in common,* (2) *the common part of every pair of cyclic elements of \bar{M} is either vacuous or itself a cyclic element of class (b), and* (3) *the set of non-internal points of a maximal cyclic curve of \bar{M} is countable.*

† H. Hahn, "Über die allgemeinste ebene Punktmenge, die stetiges Bild einer Strecke ist," *Jahresbericht der Deutschen Mathematiker Vereinigung*, vol. 23 (1914), pp. 318-322.

‡ G. T. Whyburn, "Cyclically connected continuous curves," *Proceedings of the National Academy of Sciences*, vol. 13 (1927), pp. 31-38.

Definition 14 (P. S. T., p. 115). A continuous curve is said to be *acyclic* if it contains no simple closed curve.

Definition 15 (W., § 2). A point set X will be called a *simple cyclic chain* of \bar{M} between two cyclic elements E_1 and E_2 of \bar{M} provided that X is connected, contains E_1 and E_2 , is the sum of some collection of the cyclic elements of \bar{M} , and is such that no proper connected subset of X contains E_1 and E_2 and is the sum of the elements of such a collection. The elements E_1 and E_2 are said to be the *end elements* of the chain.

LEMMA 7.† If A and B are any two points of a continuous curve \bar{M} , K denotes the set of all those points of \bar{M} which separate A from B in \bar{M} , and t is any arc in \bar{M} from A to B , then (1) $K + A + B$ is a closed set of points of t and (2) each maximal segment S of $t - (K + A + B)$ determines a unique maximal cyclic curve of \bar{M} which contains S .

Definition 16. X is a *maximal simple cyclic chain* of \bar{M} if it is not a proper subset of any other simple cyclic chain of \bar{M} .

LEMMA 8.‡ Let \bar{M} be a continuous curve in k -space and \bar{C} a simple cyclic chain of \bar{M} . Then the point set $\bar{M} - \bar{C}$ is the sum of a finite or denumerable set of components M_n , where (1) \bar{M}_n is a continuous curve, (2) $\bar{M}_n \cdot \bar{C} = \bar{M}_n - M_n =$ a single point, (3) $\bar{M}_m \cdot \bar{M}_n, m \neq n$, is null or a single point of \bar{C} , (4) every cyclic element of \bar{M} containing a point of M_n lies in \bar{M}_n , (5) every arc joining a point of \bar{M}_n to a point of $\bar{M} - M_n$ contains the point $\bar{M}_n - M_n$, and (6) for each $\epsilon > 0$, there are at most a finite number of the M_n of diameter $\geq \epsilon$.

Proof. First of all, let M_n be one of the components of $\bar{M} - \bar{C}$. If M_n contained any limit points of M_n , $M_m + M_n$ would be connected and thus $m = n$. Hence $\bar{M}_n = M_n + \bar{M}_n \cdot \bar{C}$.

Now, let M'_n be the sum of all the cyclic elements of \bar{M} which contain a point of M_n . Each point Q of $M'_n - M_n$ lies in a non-degenerate cyclic element of \bar{M} containing a point P of M_n . Obviously no such point Q can lie in $M_m, m \neq n$, since each cyclic element of \bar{M} is connected; thus every point (if any) of $M'_n - M_n$ lies in \bar{C} . Suppose Q is such a point; let c_P be the cyclic element of M'_n containing Q , and c_Q one in \bar{C} containing it. Then

† G. T. Whyburn, "Some properties of continuous curves," *Bulletin of the American Mathematical Society*, vol. 33 (1927), pp. 305-308.

‡ The results of this lemma are contained in Theorems (4.3), (4.4), (4.7), and (6.5) of Kuratowski and Whyburn, "Sur les éléments cycliques, etc.," *Fundamenta Mathematicae*, vol. 16 (1930), pp. 305-331.

$c_P \neq c_Q$ so, from Lemma 6 we conclude that $Q = c_P \cdot c_Q$, that Q is a cut point of \bar{M} and is thus the only point of $\bar{C} \cdot c_P$. Hence every point of $M'_n - M_n$ is a limit point of M_n and $\bar{M}'_n = \bar{M}_n$. This demonstrates (4).

Clearly M'_n is a connected set composed of cyclic elements of \bar{M} , so that by W., Theorems 8 and 11, \bar{M}_n is a continuous curve (which demonstrates (1)) and contains all the simple arcs of \bar{M} joining two points of \bar{M}_n . It is thus clear that $\bar{M}_n \cdot \bar{C}$ contains at most one point since \bar{C} also possesses these properties. Clearly $\bar{M}_n \cdot \bar{C} \neq 0$ since any point of M_n can be joined to a point of \bar{C} by an arc of \bar{M} which has a last point in \bar{M}_n . Thus $\bar{M}_n \cdot \bar{C} = \bar{M}_n - M_n =$ a single point. From this and the previous sentence, (2), (3), and (5) follow.

Now we have seen that any arc of \bar{M} joining a point of M_n and a point of $\bar{M} - \bar{M}_n$ must contain the point $Q_n = \bar{M}_n - M_n$. In each M_n , let P_n be as far from Q_n as possible. Now suppose there is an infinite subsequence of $\{M_n\}$, each of which is of diameter $\geq \epsilon$. Let $\{M_{n_k}\}$ be a subsequence of this so that the corresponding subsequence $\{P_{n_k}\}$ converges to P_0 . Clearly $d(P_{n_k}, Q_{n_k}) \geq \epsilon/2$ so that there exists a K such that for $k > K$, $P_{n_k} \in C(P_0, \epsilon/4)$ and $Q_{n_k} \notin C(P_0, \epsilon/4)$. Thus none of the P_{n_k} , $k > K$, can be joined to P_0 by an arc of \bar{M} lying entirely in $C(P_0, \epsilon/4)$ since each such arc must contain Q_{n_k} . But this contradicts Lemma 2.

THEOREM 1.† *Every continuous curve \bar{M} (in n -space) can be represented as the sum of a set M , described below, and a completely disconnected set M' of limit points of M , all of which are end points of \bar{M} . The set M is represented in the form $M = \bar{C}_1 + \bar{C}_2 + \dots$ where (1) each \bar{C}_n is a simple cyclic chain of \bar{M} , (2) for each n , $\bar{M}_n = \bar{C}_1 + \dots + \bar{C}_n$ is a continuous curve and $\bar{C}_{n+1} \cdot \bar{M}_n$ is a single cut point of \bar{M} which belongs to an end cyclic element of \bar{C}_{n+1} but not to an end cyclic element of \bar{C}_1 , and (3) for each $\epsilon > 0$ there exists an $N(\epsilon)$ such that, for $n > N(\epsilon)$, the diameter of each \bar{C}_n and of each component of $\bar{M} - \bar{M}_{n-1}$ is less than ϵ .*

Proof. Let d be the diameter of \bar{M} , P and Q be two points of \bar{M} such that $d(P, Q) = d$, E and F cyclic elements of \bar{M} containing P and Q respectively, and \bar{C} a maximal simple cyclic chain containing E and F such a chain being known to exist by W Theorems 13 and 15. Let $\{K_{n_1}\}$ be the sequence of components of $\bar{M} - \bar{C}$. In each K_{n_1} , let P_{n_1} be a point as far from $Q_{n_1} = K_{n_1} \cdot \bar{C}$ as possible, let E_{n_1} be a cyclic element of \bar{M} containing P_{n_1}

† The essence of this theorem is to be found in Kuratowski and Whyburn, *loc. cit.*, § 8.

(this element lies in \bar{K}_{n_1} by Lemma 8), let $F_{n_1} = Q_{n_1}$, and let \bar{C}_{n_1} be a simple cyclic chain of \bar{K}_{n_1} containing E_{n_1} and F_{n_1} as endelements. Now let $\{K_{n_1, n_2}\}$ be the components of $\bar{K}_{n_1} - \bar{C}_{n_1}$ for each n_1 ; the totality of the K_{n_1, n_2} are the components of $\bar{M} - \bar{C} - \sum \bar{C}_{n_1}$, since $K_{n_1, n_2} \subset K_{n_1}$. Also $\bar{K}_{n_1, n_2} \cdot \bar{K}_{n_1, n'_2}$ is null or a single point of \bar{C}_{n_1} , $n'_2 \neq n_2$, and $\bar{K}_{n_1, n_2} \cdot \bar{K}_{n'_1, n'_2} = 0$, $n'_1 \neq n_1$. Let $Q_{n_1, n_2} = \bar{C}_{n_1} \cdot \bar{K}_{n_1, n_2}$, let P_{n_1, n_2} be a point of \bar{K}_{n_1, n_2} as far as possible from Q_{n_1, n_2} and determine a simple cyclic chain \bar{C}_{n_1, n_2} in \bar{K}_{n_1, n_2} as above. Continue this indefinitely (i. e. unless $\bar{M} = \bar{M}_n$ for some n). If we order all of these $\bar{C}_{n_1}, \dots, n_k$ into a single sequence $\{\bar{C}_n\}$ so that $\bar{C}_n \cdot \bar{M}_{n-1} \neq 0$ and $\bar{C}_1 = \bar{M}_1 = \bar{C}$, it is clear that the conditions (1) and (2) on the set M are satisfied since no Q_{n_1}, \dots, n_k belongs to an endelement of \bar{C} and $\bar{K}_{n_1}, \dots, n_k$ has no points in common with $\bar{M} - K_{n_1}, \dots, n_k$ except a single point of $\bar{C}_{n_1}, \dots, n_{k-1}$. We shall show that, for every $\epsilon > 0$, only a finite number of all the $\bar{K}_{n_1}, \bar{K}_{n_1, n_2}, \dots, \bar{K}_{n_1}, \dots, n_k, \dots$ are of diameter $\geq \epsilon$ and this will demonstrate (3) since each $\bar{C}_{n_1}, \dots, n_k \subset \bar{K}_{n_1}, \dots, n_k$ and each component of $\bar{M} - \bar{M}_n$ is a K_{n_1}, \dots, n_k for some k . Then it will be clear (using (3)) that every point of $\bar{M} - M$ is a limit point of M , and hence, by W., Theorem 11, is an end point of \bar{M} . By P. S. T., chapter II, Theorem 42, this set is completely disconnected.

Now suppose there were an infinite number of the \bar{K} 's which were of diameter $\geq \epsilon > 0$. By repeated use of the preceding lemma we know that, for each k , only a finite number of the $\bar{K}_{n_1}, \dots, n_j$, $1 \leq j \leq k$, are of diameter $\geq \epsilon$. Hence it is easy to see that we can find a sequence $\bar{K}_{n_1} \supset \bar{K}_{n_1, n_2} \supset \dots$, all of which are of diameter $\geq \epsilon$, each $\bar{K}_{n_1}, \dots, n_k$ being a particular one of the $\bar{K}_{n_1}, \dots, n_k$. Let $\bar{K} = \prod_{k=1}^{\infty} \bar{K}_{n_1}, \dots, n_k$; \bar{K} is of diameter $\geq \epsilon$. Now each $\bar{C}_{n_1}, \dots, n_k$ is so chosen that it contains a point P_{n_1}, \dots, n_k as far as possible and hence at a distance $\geq \epsilon/2$ from $\bar{Q}_{n_1}, \dots, n_k = \bar{C}_{n_1}, \dots, n_k \cdot \bar{C}_{n_1}, \dots, n_{k-1}$. We have seen that any arc joining a point of $\bar{K}_{n_1}, \dots, n_k$ to a point of $\bar{C}_{n_1}, \dots, n_{k-1}$ must contain Q_{n_1}, \dots, n_k . Hence it is easy to see that any arc joining a point of \bar{K} to a point of $\bar{C}_{n_1}, \dots, n_k$ must contain all the Q_{n_1}, \dots, n_l , $l > k$. Then let $\{l_k\}$ be a subsequence of the integers such that

$$\lim_{k \rightarrow \infty} P_{n_1}, \dots, n_{l_k} = P_0 \quad \text{and} \quad \lim_{k \rightarrow \infty} Q_{n_1}, \dots, n_{l_k} = Q_0.$$

Obviously $P_0 \in \bar{K}$, $Q_0 \in \bar{K}$, and $d(P_0, Q_0) \geq \epsilon/2$. Since every arc joining P_{n_1}, \dots, n_{l_k} to P_0 must go through all the points Q_{n_1}, \dots, n_{l_m} , $m > k$, it must go through Q_0 . But this contradicts Lemma 2.

2. *Hemicactoids and correspondences.* In this section, we shall intro-

duce the notion of a "hemicactoid" and bring out the relationship between hemicactoids and "upper semicontinuous collections of mutually exclusive continua filling up" a plane Jordan region. The connection between the theory of these collections and the theory of surfaces is suggested by Theorem 1, below.

Definition 1 (P. S. T., p. 28). We shall say that a continuum \bar{C} , is the *limit continuum* of the sequence $\{\bar{C}_n\}$ of continua if

- (i) all the limit points of a sequence $\{P_n\}$ of points, $P_n \in \bar{C}_n$, lie in \bar{C} , and
- (ii) if P is any point of \bar{C} , there exists a sequence $\{P_k\}$, $P_k \in \bar{C}_{n_k}$, of points which converges to P , $\{n_k\}$ being a subsequence of the integers.

LEMMA 1 (P. S. T., chapter I, Theorem 42). If $\{\bar{C}_n\}$ is a sequence of continua such that there exists a convergent sequence $\{P_n\}$ of points, $P_n \in \bar{C}_n$, then the sequence $\{\bar{C}_n\}$ possesses a unique limit continuum \bar{C} , which consists of all points P satisfying (ii), Definition 1.

Definition 2 (cf. P. S. T., chapter V, p. 324). A collection G of continua is *upper semicontinuous* provided it is true that if g is a continuum of G and $\{g_n\}$ is a sequence of continua of G , each g_n containing a point p_n such that the sequence $\{p_n\}$ converges to a point p of g , then g contains the limit continuum of the g_n .

Definition 3. Given an upper semicontinuous collection G of continua, a continuum g of G , and a sequence $\{g_n\}$ of continua of G . g will be said to be the *unique limit continuum with respect to G* of the sequence $\{g_n\}$ if and only if every limit point of any sequence $\{p_n\}$ of points where $p_n \in g_n$ for each n , lies in g . If there is a sequence $\{p_n\}$ of points, where $p_n \in g_n$ which converges to a point p of g , it is clear from Definition 2 that g will be the *unique limit element* of $\{g_n\}$ (with respect to G).

Remarks. Using P. S. T., chapter V, Theorems 1 and 2, we see that this definition is equivalent to that given in P. S. T., page 326. Clearly, if the continua of a collection are all points, this notion of limit element reduces to the ordinary notion of limit point. It is also clear that the notions of closed set and connected set of elements of G may be defined in terms of limit element.

LEMMA 2 (P. S. T., chapter V, Theorems 4 and 5). A necessary and sufficient condition that a set H , of elements of an upper semicontinuous collection of mutually exclusive continua be closed or connected is that the set of points covered by H be closed or connected, respectively.

Definition 4. A correspondence between the elements of two upper semicontinuous collections of mutually exclusive continua will be said to be *topological* if it is 1 — 1 and limit elements are preserved.

Remark. It is clear that if T_1 is a topological transformation of G_0 into G_1 and T_2 of G_1 into G_2 , then $T_1 \cdot T_2$ is a topological transformation of G_0 into G_2 if the G_i are all upper semicontinuous collections of mutually exclusive continua (which may all reduce to points in a given collection).

Definition 5. A vector function, $U = \phi(u)$, defined on a continuum \bar{C} , is said to be *monotone* if it is continuous and for every U_0 , the set c_0 of points u for which $\phi(u) = U_0$ is null or a continuum.

THEOREM 1. *If $\phi(u)$ is a continuous vector function defined over a connected set C which is such that the sets c_0 of points u of C for which $\phi(u) = \phi(u_0)$ are all closed, then the maximal continua over each of which $\phi(u)$ is constant form an upper semicontinuous collection of mutually exclusive continua filling up C .*

Proof. Let $\{c_n\}$ be a sequence of the above continua such that there exists a sequence $\{p_n\}$ of points, p_n being in c_n for each n , which converges to a point p of C and let c be the continuum of the collection which contains p . Let \bar{c} be the limit continuum of $\{c_n\}$ and let $q \in \bar{c}$; $p \in \bar{c}$ of course. There exists a sequence $\{q_{n_k}\}$, $q_{n_k} \in c_{n_k}$ converging to q . Then

$$\phi(q) = \lim_{k \rightarrow \infty} \phi(q_{n_k}) = \lim_{k \rightarrow \infty} \phi(p_{n_k}) = \phi(p),$$

and thus $\phi(u)$ is constant over \bar{c} . Hence $\bar{c} \subset c$ and thus the given collection is upper semicontinuous.

THEOREM 2. *Let G be any upper semicontinuous collection of continua filling up a connected set C which is the outer limit of a sequence of continua and let T be a transformation of the elements of G into the points of a set K . Let $U(u)$ be the vector function defined on C by the condition that $U(u) = U_P$ when u belongs to a continuum of G which is carried by T into the point P . Then a necessary and sufficient condition that T be topological is (1) that $U(u)$ be monotone and (2) the collection of maximal continua over each of which $U(u)$ is constant is identical with G .*

Proof. Suppose T is topological. Let $\{u_n\}$ be a sequence of points of C converging to the point u_0 of C , let c_0 be the continuum of G containing u_0 , and let c_n be that containing u_n , for each n . From the definition of limit element of G , it follows that c_0 is the unique limit element of $\{c_n\}$. Hence $U(u_0) = \lim_{n \rightarrow \infty} U(u_n)$ and $U(u)$ is thus continuous. Furthermore, since T is

1 — 1, every point u of C for which $U(u) = U(u_0)$ lies in c_0 . Hence $U(u)$ is monotone and (2) is satisfied.

Now suppose conditions (1) and (2) are satisfied. Let C be the outer limit of the continua \bar{C}'_m , let $\bar{C}'_m \subset \bar{C}'_{m+1}$ for each m , and, for each m , let G_m be the subcollection of G consisting of those continua of G which have a point in common with \bar{C}'_m . Clearly the set \bar{C}_m of points covered by G_m is a continuum since it is obviously connected and is closed by P. S. T., chapter V, Theorem 2. Furthermore, G_m is upper semicontinuous by P. S. T., chapter V, Theorem 1. Also (1) C is the outer limit of the continua \bar{C}_m , (2) the set of points \bar{K}_m into which \bar{C}_m is carried by $U = U(u)$ is a continuum for each m and (3) K is the outer limit of the \bar{K}_m . Hence if we show that $T, T : U = U(u)$, yields a topological transformation of G_m into \bar{K}_m for each m , the desired result will follow.

Clearly T is the transformation of the maximal continua of \bar{C}_m over each of which $U(u)$ is constant into the points of \bar{K}_m and is therefore 1 — 1, $U(u)$ being monotone. The continuity of T from G_m to \bar{K}_m is immediate. Now let $\{P_n\}$ be a sequence of points of \bar{K}_m converging to the point P_0 (of \bar{K}_m), let c_0 be the continuum of G_m corresponding to P_0 , and for each n let c_n be that corresponding to P_n . Let $\{p_{n_k}\}$ be a convergent sequence of points, $p_{n_k} \in c_{n_k}$. The limit point p_0 of this sequence is in \bar{C}_m and, since $U(u)$ is continuous, $U(u_{p_0}) = \lim_{k \rightarrow \infty} U(u_{p_{n_k}}) = U_{p_0}$, so that $p_0 \in c_0$. Thus c_0 is the limit continuum of $\{c_n\}$ in G_m and hence T is continuous.

LEMMA 3. *A necessary and sufficient condition that a plane continuum should fail to separate the plane is that it be representable as the product of a sequence $\{\bar{R}_n\}$ of closed Jordan regions each of which contains the next.*

Proof. The condition is obviously sufficient.

Now suppose that the continuum \bar{C} does not separate the plane Π . Then any point P of $\Pi - \bar{C}$ can be joined to ∞ by an arc, all of whose points are at a distance $\geq d_P > 0$ from \bar{C} . Let \bar{R}_n be the set of all points bounded by the outer simple closed curve of the set \bar{R}'_n of points at a distance $\leq 1/n$ from \bar{C} ; \bar{R}'_n is known to be a region bounded by a finite number of simple closed curves. Then $\bar{C} = \bigcap_{n=1}^{\infty} \bar{R}_n$, for any point P in \bar{C} is in each \bar{R}_n and any point P in $\Pi - \bar{C}$ is outside of \bar{R}_n for $n > 1/d_P$.

Definition 6. By a *base set* we shall mean a bounded continuous curve in the plane which does not separate the plane.

LEMMA 4. *Let G be an upper semicontinuous collection of mutually exclusive continua filling up a Jordan region \bar{r} , no one of which separates the*

plane although some may separate \bar{r} . Then G is topologically equivalent to a base set.

Proof. Let G_1 be the collection, filling up the plane Π , consisting of G plus all the points of Π outside \bar{r} . Clearly G_1 is upper semicontinuous and no continuum of G_1 separates Π . From a theorem of R. L. Moore,[†] we can find a 1 — 1 continuous mapping $T, T : U = U(u)$, $U(u)$ monotone, of the continua of G_1 on to the points of Π . Let $\{r_n^*\}$ be a sequence of simple closed curves approaching r^* uniformly, each containing the next and r^* in its interior. Now evidently r_n^* is carried in a 1 — 1 continuous way into a Jordan curve R_n^* for each n . Now no point interior to r_n^* corresponds to a point exterior to R_n^* for first, we can find points interior and near to r_n^* which are carried interior to R_n^* and second, the set of points corresponding to r_n is connected. Similarly, points outside r_n^* go into points outside R_n^* and thus T carries all of r_n into all of R_n^* and all of $\Pi - \bar{r}_n$ into all of $\Pi - \bar{R}_n$, \bar{R}_n being the Jordan region bounded by R_n^* . Hence each \bar{R}_n contains the next, all contain $T(\bar{r})$, and any point not in $T(\bar{r})$ is outside of \bar{R}_n for some n so that $T(\bar{r}) = \bigcap_{n=1}^{\infty} \bar{R}_n$. That $T(\bar{r})$ is a continuous curve follows from the fact that it is the range of values of the continuous vector function $U(u)$ (see Lemma 3, § 1).

Definition 7. By a *double cone*, we shall mean the surface of revolution generated by revolving an isosceles triangle about its base as axis. By the *axis* of the double cone, we mean its axis of revolution; this segment is not part of the surface. By the *end points* of the double cone, we mean the end points of its axis, and by its *vertex angle*, the magnitude of a base angle of a generating isosceles triangle.

Definition 8. Given a rhombus, we shall designate two opposite vertices as "*end points*," define its *vertex angle* as half the angle of the rhombus subtended at one of the end points, and define its *axis* as the diagonal segment joining its end points.

Definition 9. By a *simple cyclic chain of type A (type B)* we shall mean a point set obtained as follows: let L be a closed line segment and F a closed set on it; replace each complementary interval of F on L by a double cone (rhombus) with the same end points, the vertex angles of the double cones (rhombuses) being the same.

Definition 10 (P. S. T., p. 151). By a *cactoid*, we mean a bounded con-

[†] R. L. Moore, "Concerning upper semicontinuous collections of continua," *Transactions of the American Mathematical Society*, vol. 27 (1925), p. 425.

tinuous curve (in a metric space) all of whose non-degenerate simple links are simple closed surfaces.

Definition 11. By a *canonical cactoid (base set)*, we mean a bounded continuous curve in 3-space which may be expressed in the form in Theorem 1, § 1, the \bar{C}_n being all simple cyclic chains of type A (type B).

LEMMA 5. Any simple cyclic chain of a cactoid (base set) can be mapped in a 1—1 continuous way on a simple cyclic chain of type A (type B).

Proof. From the definition of cactoid and from the fact (Lemma 4, § 1) that every non-degenerate simple link of a continuous curve \bar{M} is a non-degenerate cyclic element of \bar{M} and conversely, we see that every maximal cyclic curve of a cactoid is a simple closed surface. From W., Theorem 24, we see that a necessary and sufficient condition that a plane continuous curve fail to separate the plane is that each of its maximal cyclic curves is a closed Jordan region.

Now let \bar{C} be the simple cyclic chain under consideration and let E_1 and E_2 be its endelements. Let $P_i = E_i$ if E_i is a point or let P_i be an internal point (Definition 12, § 1) of E_i if E_i is non-degenerate, ($i = 1, 2$), and let t be an arc joining P_1 and P_2 . Let t be mapped in a 1—1 continuous way on a line segment L of the desired length. Let K be the set of points separating P_1 from P_2 in \bar{C} ; by Lemma 7, § 1, all the points of K lie on t and $K + P_1 + P_2$ is closed. Let Σ be the closed set on L corresponding to $K + P_1 + P_2$. By Lemma 7, § 1 and W., Theorem 1, each arc of t corresponding to a complementary interval of Σ on L determines a unique maximal cyclic curve of \bar{C} and \bar{C} is the sum of $K + E_1 + E_2$ and all of these maximal cyclic curves. Furthermore, if an E_i is non-degenerate, there is an arc of t , with P_i as end point, lying in E_i in which case, we say that E_i is determined by this arc. Now map each of these maximal cyclic curves of \bar{C} corresponding to a given arc of t on the double cone (if \bar{C} belongs to a cactoid) or rhombus (if \bar{C} belongs to a base set) which has the corresponding end points on L and the given vertex angle. Clearly this 1—1 map of \bar{C} on $\bar{\Gamma}$, the canonical chain, is continuous for, by the argument of Lemma 8, § 1, there are only a finite number of maximal cyclic curves of \bar{C} having a diameter $\geq \epsilon$, for each $\epsilon > 0$.

Definition 12. A *hemicactoid* is a continuous curve in 3-space which is the sum of a canonical base set \bar{B} , in the (x, y) plane Π and a finite or denumerable set of canonical cactoids \bar{K}_n , each one of which has exactly one point in common with \bar{B} , is otherwise entirely above Π , and has no point, not in \bar{B} , in common with any of the other cactoids.

Definition 13. Two hemicactoids are said to be *homeomorphic* if and

only if they can be put into a 1 — 1 continuous correspondence in which the base sets are also in a 1 — 1 continuous correspondence.

The following lemma is immediate:

LEMMA 6. Let \bar{M} be a hemicactoid $\bar{M} = \bar{B} + \sum_{m=1}^{(\infty)} \bar{K}_m + M'$; (the symbol (∞) means that the upper limit may be finite or infinite) the number of simple cyclic chains in either \bar{B} or the totality of the \bar{K}_m being finite. Let (1) $\epsilon > 0$, (2) $\eta > 0$, (3) $\bar{\Gamma}$ be a simple cyclic chain of type A if the totality of the \bar{K}_m contains only a finite number of chains or of type B if this is not the case, and (4) P be a point of one of the \bar{K}_m not on an end cyclic element other than $\bar{K}_m \cdot \bar{B}$ of a \bar{K}_m in the first case above, or any point of B^* not on an end cyclic element of \bar{B}_1 (one chain of \bar{B}) in the second case. Then we can find a canonical (i. e. of type A or B) simple cyclic chain \bar{C} , homeomorphic to $\bar{\Gamma}$, which may be attached to \bar{M} at P in such a way that $\bar{M} + \bar{C}$ is a hemicactoid and, for $\rho \geq \epsilon$, $\eta_{\bar{M}+\bar{C}}(\rho) > (1 - \eta)\eta_{\bar{M}}(\rho)$, and for $0 < \rho < \epsilon$, $\eta_{\bar{M}+\bar{C}}(\rho) > 0$, where $\eta_{\bar{L}}(\rho)$ is the lower boundary, for $d_{\bar{L}}(P, Q) \geq \rho$ of the ratio of $d(P, Q)$ to $d_{\bar{L}}(P, Q)$, $d_{\bar{L}}(P, Q)$ being the geodesic distance along \bar{L} from P to Q .

LEMMA 7. Any cactoid (base set) can be mapped in a 1 — 1 continuous way on a canonical cactoid (base set).

Proof. Let \bar{M}' be the given cactoid (base set) and let it be represented as in Theorem 1, § 1. By Lemma 5, each \bar{C}'_n can be mapped in a 1 — 1 continuous way on a simple cyclic chain of type A (type B). Let \bar{C}'_1 be so mapped on such a chain \bar{C}_1 of unit length and vertex angle $\leq \pi/4$. By Lemmas 5 and 6, we can find a simple cyclic chain \bar{C}_2 , homeomorphic to \bar{C}'_2 , of length $\leq 1/4$ and vertex angle $\leq \pi/4$ which can be attached to \bar{C}_1 at the point of \bar{C}_1 (not on an endelement of \bar{C}_1) corresponding to the point $\bar{C}'_1 \cdot \bar{C}'_2$ in such a way that $\bar{C}_1 + \bar{C}_2$ is a canonical cactoid (base set) and $\eta_2(\rho) = \eta_{\bar{C}_1+\bar{C}_2}(\rho) \geq (1 - 1/2^2)\eta_1(\rho)$ for $\rho \geq 1/2$, and $\eta_2(\rho) > 0$ for $0 < \rho < 1/2$, $\eta_1(\rho)$ being $\eta_{\bar{C}_1}(\rho)$. It is clear from Lemma 6 that this process may be continued indefinitely with $\epsilon = 1/n$, $\eta = 1/n^2$, the length of $\bar{C}_n \leq 1/n^2$, and the vertex angle of $\bar{C}_n \leq \pi/4$. At each stage, the correspondence between \bar{M}'_n and \bar{M}_n set up by uniting these individual correspondences between the chains is 1 — 1 and uniformly continuous both ways. Also these unite to give a 1 — 1 continuous correspondence between \bar{M} and \bar{M}' (cf. Theorem 1, § 1). Furthermore it is clear that there exists a continuous monotone function $\eta(\rho)$, positive for $\rho > 0$ and vanishing with ρ , such that $\eta(\rho) \leq \eta_n(\rho)$ for each n . We shall show † that this correspondence is uniformly continuous both ways

† The argument from this point on is essentially that used by H. M. Gehman in his

so that it can be extended to the closures, \bar{M} and \bar{M}' of M and M' respectively which shows also that \bar{M} is a continuous curve and thus a canonical cactoid (base set).

Suppose the mapping of M onto M' is not uniformly continuous. Then there exist sequences $\{P_i\}$ and $\{Q_i\}$ of points of M and an $\epsilon > 0$ such that $d(P_i, Q_i) < 1/i$ and $d(P'_i, Q'_i) \geq \epsilon$. Let n be so large that the diameter of any component of $M' - \bar{M}'_n < \epsilon/3$. Then any arc of M' joining P'_i and Q'_i contains an arc $\widehat{E'_i F'_i}$ in \bar{M}_n (entirely, by W., Theorem 8) of diameter $> \epsilon/3$. On the other hand there exists a $\delta > 0$ such that if $d(A, B) < \delta$, $A, B \in \bar{M}_n$, then $d(A', B') < \epsilon/3$. Also, since M has the property that any two of its points at a distance $\leq \rho$ apart can be joined by an arc of M of length $\leq \eta^{-1}(\rho)$, $\eta^{-1}(\rho)$ being the inverse function of $\eta(\rho)$ and approaching zero with ρ , we see by W., Theorem 9, that M is uniformly connected in kleinen and there exists an $\alpha > 0$ such that if $d(A, B) < \alpha$, $A, B \in M$, A and B can be joined by an arc of M of diameter $< \delta$. Hence if $i > 1/\alpha$, $d(P_i, Q_i) < \alpha$ and they can be joined in M by an arc of diameter $< \delta$ so that the arc $\widehat{E_i F_i}$ in \bar{M}_n is of diameter $< \delta$ and thus the diameter of $\widehat{E'_i F'_i} < \epsilon/3$, contrary to the third sentence of this paragraph.

Now, it is clear that, for every $\epsilon > 0$, there exists an $N(\epsilon)$ such that, for $n > N(\epsilon)$, the diameter of each maximal connected subset of $M - \bar{M}_n < \epsilon/3$. Also, by Lemma 2, § 1, \bar{M}' is uniformly arcwise connected in kleinen. Hence the above argument can be used to demonstrate the uniform continuity of the mapping of M' on \bar{M} . This is what we wished to prove.

LEMMA 8. *Let G be an upper semicontinuous collection of mutually exclusive continua filling up \bar{r} and let G_0 be the subcollection of G consisting of the continua g_0 of G such that no point of g_0 is in a bounded complementary domain \dagger of any continuum of G . Then every point P of \bar{r} is either in a g_0 or in a bounded complementary domain of some g_0 .*

Proof. Let G be the given collection and G_0 the above subcollection. It is clear that if one point of a continuum g of G is in a complementary domain of another continuum of G , all of g is in that complementary domain. Now a given point P of \bar{r} is either in a g_0 or is in a bounded complementary domain of some continuum $g^{(1)}$ of G ; either $g^{(1)} \in G_0$ or it lies wholly in a

paper "Concerning acyclic continuous curves," *Transactions of the American Mathematical Society*, vol. 29 (1927), pp. 553-555 in particular.

\dagger A complementary domain of a continuum \bar{C} in the plane Π is a component of $\Pi - \bar{C}$. A bounded complementary domain of a continuum is an open simply connected region.

complementary domain of a continuum $g^{(2)}$ of G which also contains a circular disc outside of $g^{(1)}$. By a repetition of this argument, we obtain a normally ordered set $g^{(1)}, g^{(2)}, \dots, g^{(\omega)} g^{(\omega+1)}, \dots$, each $g^{(\alpha)} \in G$ and being wholly inside a bounded complementary domain of $g^{(\alpha+1)}$ which also contains a circular disc outside of $g^{(\alpha)}$. Thus this set is denumerable and hence there is a first ordinal β in Cantor's second class for which $g^{(\beta)} \in G_0$. This proves the lemma.

THEOREM 3. *Any upper semicontinuous collection of mutually exclusive continua filling up a Jordan region \bar{r} is topologically equivalent to a hemi-cactoid. We may further require that a monotone transformation $U = U(u)$ of \bar{r} into \bar{H} carry the subcollection G_0 of Lemma 8 into the base set \bar{B} . If $U = U(u)$ and $U = U_1(u)$ are two such transformations of \bar{r} into \bar{H} and \bar{H}_1 respectively, then there exists a 1—1 continuous transformation $U_1 = \bar{U}_1(U)$ of \bar{H} into \bar{H}_1 which carries \bar{B} into \bar{B}_1 topologically and is such that $\bar{U}_1[U(u)] = U_1(u)$.*

Proof. Now let G be the given collection and G_0 the subcollection of Lemma 8. If we form the collection G'_0 of continua g'_0 , where each g'_0 is a g_0 plus all of its bounded complementary domains, we see that no continuum of G'_0 separates Π , G'_0 is upper semicontinuous and topologically equivalent to G_0 and, by Lemma 8, fills up \bar{r} . Thus G_0 is topologically equivalent to a canonical base set \bar{B} . Let $U = U(u)$ yield such a correspondence.

Now the subcollection of G which lies in a bounded complementary domain d of some g_0 plus that g_0 is topologically equivalent to an upper semicontinuous collection filling up the surface of a sphere, for there exists a monotone transformation of \bar{d} into the sphere which carries d^* into a point and is otherwise 1—1. Such a collection is known, from a theorem of R. L. Moore (P. S. T., Theorem 20, chapter VII) and from Lemma 7, to be topologically equivalent to a canonical cactoid. It is clear that there are at most a denumerable infinity of such bounded complementary domains, so let them be ordered into a single sequence $\{d_m\}$ and let \bar{K}'_n be a canonical cactoid corresponding to $d_n + g_0^{(n)}$ ($g_0^{(n)}$ being the g_0 containing d_n^*), the correspondence being given by $U = U'_n(u)$. Let $\bar{K}'_n = \sum_{p=1}^{\infty} \bar{C}'_{n,p} + K'^*_n$, and let the $\bar{C}'_{n,p}$ be ordered into a single sequence $\{\bar{C}_m\}$ such that

$$\bar{C}_{m+1} \cdot (\bar{B} + \bar{C}_1 + \dots + \bar{C}_m) \neq 0$$

for any $m \geq 0$. Now by Lemma 6, we can find a simple cyclic chain \bar{C}_1 , of type A, homeomorphic to \bar{C}_1 and of length ≤ 1 , and can join it onto \bar{B} , projecting upwards, at the point of \bar{B} corresponding to the g_0 containing the

boundary of the \bar{d}_a in \bar{r} corresponding to the cactoid containing \bar{C}_1 . Clearly $\bar{B} + \bar{C}_1$ is a hemicactoid. Let $U(u)$ be extended to the part \bar{d}_{1a} of \bar{d}_a which corresponds to \bar{C}_1 so that $U = U(u)$ gives a topological correspondence between $G_0 + G \cdot \bar{d}_{1a}$ and $\bar{B} + \bar{C}_1$. By Lemma 6, we can find a \bar{C}_2 homeomorphic to \bar{C}_1 of length $\leq 1/4$, which can be attached to $\bar{B} + \bar{C}_1$ at the point corresponding to the continuum of $G_0 + G \cdot \bar{d}_{1a}$ which bounds the complementary domain of $\hat{G}_0 + \bar{d}_{1a}$ (\hat{G}_0 being the set of points covered by G_0) in which the continua corresponding to \bar{C}_2 lie,† and is such that

$$\begin{aligned} \eta_{\bar{B} + \bar{C}_1 + \bar{C}_2}(\rho) &= \eta_2(\rho) \geq (1 - 1/2^2)\eta_1(\rho), & \rho &\geq 1/2, \\ \eta_2(\rho) &> 0, & 0 < \rho < 1/2, \end{aligned}$$

where $\eta_1(\rho) = \eta_{\bar{B} + \bar{C}_1}(\rho)$. From Lemma 6, it is clear that this may be continued indefinitely so that the length of $\bar{C}_n \leq 1/n^2$ and $\eta_n(\rho) \geq [1 - 1/n^2]\eta_{n-1}(\rho)$ for $\rho \geq 1/n$ and $\eta_n(\rho) > 0$ for $0 < \rho < 1/n$. By the method of Lemma 7 it can be shown that the \bar{C}_i corresponding to the \bar{C}_i of one of the \bar{K}'_n plus the limit points of this set not in any \bar{C}_i add up to a canonical cactoid \bar{K}_n homeomorphic to \bar{K}'_n , that the closure \bar{H} of the set $H = \bar{B} + \sum_{i=1}^{\infty} \bar{C}_i$ is a hemicactoid, that $\bar{H} = \bar{B} + \sum_{n=1}^{\infty} \bar{K}_n$, and that $U(u)$ extended as indicated to all of \bar{r} remains continuous and yields a topological transformation of G into \bar{H} and G_0 into \bar{B} , being monotone consequently.

The remaining statements are immediate, remembering the fact (remark after definition 4) that the product of two topological transformations of upper semicontinuous collections of continua is again topological and that the notions of limit point and limit element in a collection of continua are equivalent if the continua are all points (remarks after definition 3).

LEMMA 9. *Given a base set (cactoid) in n -space. There exists an upper semicontinuous collection of mutually exclusive continua filling up a Jordan region \bar{r} (the surface Σ of a sphere) which is topologically equivalent to the given point set. In the case of the base set it may be further required that no continuum of the corresponding collection separates Π and also that every non-degenerate continuum separates \bar{r} . In the case of the cactoid, it may be assumed that each non-degenerate continuum of the collection separates Σ .*

Proof. The lemma for cactoids has already been proved by R. L. Moore

† That all of the continua corresponding to \bar{C}_2 lie in such a bounded complementary domain follows from the arguments of Lemma 8.

(see P. S. T., chapter VII, Theorem 23). We shall give a brief proof of this part of the lemma using the result for base sets, which we shall now prove.

Let \bar{B} ($\bar{B} = B' + \sum_{i=1}^{\infty} \bar{C}_i$, where the \bar{C}_i are simple cyclic chains of type B , $\bar{C}_i \cdot \bar{C}_j$ is at most one point, $i \neq j$, and B' is completely disconnected as in Theorem 1, § 1) be a canonical base set in the plane Π , homeomorphic to the given base set. Let $\Pi - \bar{B}$ be mapped conformally on $\Pi - \bar{R}$, \bar{R} being the the closed unit circle (clearly there is no loss of generality in taking $\bar{r} = \bar{R}$), by the vector function $U = U(u)$, $u \in \Pi - \bar{R}$. Now, by Lemma 1, § 3, $U(u)$ remains continuous on R^* and is not constant over any continuum of R^* . We shall proceed to extend $U(u)$ to all of \bar{R} so that it is monotone, and $U = U(u)$ carries \bar{R} into \bar{B} in such a way that any non-degenerate continuum over which $U(u)$ is constant separates \bar{R} (but not Π) and is carried by $U = U(u)$ into a cut point of \bar{B} .

Let Σ_1 be the closed sub-set of R^* which is carried into $(\Pi - \bar{C}_1)^*$. It is clear from the well known nature of $U(u)$ on $(\Pi - \bar{R})^*$, that the end points of each complementary interval of Σ_1 correspond to the same point of $(\Pi - \bar{C}_1)^*$. Hence join each of these pairs of points by a segment and define $U(u)$ to be constant over each such segment; let R_1 be the convex region bounded by the simple closed curve formed by these segments and the points of R^* not cut off by any of them. If we now define the elements of R_1^* to be the continua of R_1^* over which $U(u)$ is constant, it is seen that at most two of these elements are carried into the same point of $(\Pi - \bar{C}_1)^*$. Let each point of each element of a corresponding pair be joined to each point of the corresponding element by a segment, and define $U(u)$ to be constant over each such segment. Now, the remainder of R_1 consists of Jordan regions ρ_1 such that $U = U(u)$ gives a representation of the boundary of a rhombus of \bar{C}_1 on the boundary ρ_1^* ; hence define $U(u)$ to be continuous in each such ρ_1 and such that $U = U(u)$ gives a 1 — 1 continuous map of the interior of the rhombus on the interior of ρ_1 .

Now, a unique one of the segments on the boundary of R_1 is carried by $U = U(u)$, so far defined, into the unique point $\bar{C}_1 \cdot \bar{C}_2$ and has the additional property that it cuts off all the points of R^* carried into points of $\bar{C}_2 - \bar{C}_1 \cdot \bar{C}_2$. Let Σ_2 be this segment plus the set of all these points of R^* . Again, each pair of end points of a complementary interval of Σ_2 corresponds to the same point of \bar{C}_2 . Join all these pairs of points by line segments and define $U(u)$ to be constant over each such segment. Let R_2 be the convex region bounded by all these new segments and Σ_2 . If each continuum of R_2^* over which $U(u)$ is constant be considered as an element, it is again seen that at most two such elements are carried by $U = U(u)$ into the same point of \bar{C}_2 . Join each point

of each element of a pair of corresponding elements to each point of the corresponding element by a segment and define $U(u)$ to be constant over each such segment. Clearly the rhombuses of \bar{C}_2 may be mapped in a 1 — 1 continuous way on the remaining regions of R_2 , as before.

It is now clear that this may be continued indefinitely and that finally $U(u)$ will be defined over all of \bar{R} . It is clear that it will be continuous, since, for every $\epsilon > 0$, at most a finite number of the \bar{C}_n and a finite number of the rhombuses are of diameter $\geq \epsilon$. It is also immediate that it is monotone and that every continuum of \bar{R} which is carried into a point of \bar{B} separates \bar{R} and is carried into a cut point of \bar{B} .

To prove the lemma for cactoids, we first replace the cactoid by a homeomorphic canonical one \bar{C} , the axes of whose simple cyclic chains form an acyclic continuous curve lying entirely in Π (that this can be done follows from Lemmas 5 and 6 and the argument of Lemma 7). Now, we map the closed upper half of \bar{C} on the closed upper half of a hemisphere by projecting it onto Π , forming a canonical base set \bar{B} , mapping \bar{B} on \bar{R} , and then projecting up onto the hemisphere from \bar{R} . We then extend the mapping by making a point P on the lower half of the cactoid to correspond to that point on the lower hemisphere which is the reflection in Π of the point of the upper hemisphere corresponding to the reflection of P on the cactoid.

THEOREM 4. *Given any hemicactoid \bar{H} , we can find a monotone vector function $U(u)$ such that $U = U(u)$ carries \bar{r} into \bar{H} . If G is the upper semicontinuous collection of continua over each of which $U(u)$ is constant, i. e., the collection corresponding to $U(u)$, and G_0 is the subcollection of Lemma 8, we may require that $U = U(u)$ carry G_0 into \bar{B} .*

Proof. By Lemma 9, we can find an upper semicontinuous collection $G_0^{(0)}$ filling up \bar{r} which is topologically equivalent to \bar{B} and such that none of its continua separate the plane. Let $U = U_0(u)$ be a monotone transformation giving such a correspondence. Now order the cactoids of \bar{H} having only one point each in common with \bar{B} into a sequence $\{\bar{K}_n\}$. Let us take $\bar{r} = \bar{Q}$ (obviously we may take $\bar{r} = \bar{Q}$ without loss of generality). Now \bar{K}_1 joins on to \bar{B} at a point P_1 ; let p_1 be a point of the continuum of $G_0^{(0)}$ corresponding to P_1 . Make a transformation of \bar{Q} into itself defined as follows: each point $p \neq p_1$ shall correspond to the point p' on the directed ray p_1p such that $d(p', p'') = d(p, p'')/2$, p'' being the last point of \bar{Q} on the ray p_1p ; the point p_1 will be carried into all the points p' on the boundary of the square $d(p_1, p') = d(p_1, p'')/2$, p'' being the last point of \bar{Q} on the ray p_1p' . Now define $U_1(u)$ outside this square as the transform of $U_0(u)$ and define it inside and on this square so as to be continuous in \bar{Q} and carry

the closed square in a monotone way into \bar{K}_1 . That this can be done follows from the fact that we may carry it in a monotone way onto the surface of a sphere \mathfrak{S} , the boundary being carried into a point and then, by Lemma 9, we can carry \mathfrak{S} in a monotone way into \bar{K}_1 . $U_1(u)$ is clearly monotone.

Now \bar{C}_2 is attached to \bar{B} at a point P_2 (perhaps $= P_1$); let p_2 be a point corresponding under $U = U_1(u)$ to P_2 . Make a transformation of \bar{Q} into itself similar to the above in which any point $p \neq p_2$ is carried into the point p' on the ray p_2p such that $d(p', p'') = (1 - 1/2^2)d(p, p'')$ and p_2 is carried into all the points p' satisfying $d(p_2, p') = d(p_2, p'')/2^2$, p'' being the last point of \bar{Q} on p_2p' . Define $U_2(u)$ outside this square as the transform of $U_1(u)$ and within and on the square so as to be continuous in \bar{Q} and carry the closed square in a monotone way into \bar{K}_2 . Clearly $U_2(u)$ is monotone and $U = U_2(u)$ carries \bar{Q} into $\bar{B} + \bar{K}_1 + \bar{K}_2$ and the twice transformed collection $G_0^{(2)}$ topologically into \bar{B} .

Let this be continued indefinitely. A simple analysis serves to show that the sequence $\{U_n(u)\}$ converges uniformly to a monotone function $U(u)$ with corresponding collection G , that $G_0 = G_0^{(\infty)}$, and that $U = U(u)$ carries \bar{Q} into \bar{H} and G_0 topologically into \bar{B} .

THEOREM 5. *Let \bar{H} be a hemicactoid with base set \bar{B} and canonical cactoids \bar{C}_m ; we have $\bar{B} = \sum_{n=1}^{\infty} \bar{B}_n + B'$ and $\bar{C}_m = \sum_{n=1}^{\infty} \bar{C}_{m,n} + C'_m$, where the B' and C'_m are mutually exclusive completely disconnected sets and \bar{B}_n and $\bar{C}_{m,n}$ are simple cyclic chains of types B and A respectively joined together as in Definition 12 and Theorem 1, § 1. Now on each \bar{B}_n and $\bar{C}_{m,n}$, let there be determined a continuous monotone transformation $T_n^{(2)}$ (for \bar{B}_n) and $T_{m,n}^{(1)}$ (for $\bar{C}_{m,n}$) of each into itself, which carries each non-degenerate cyclic element into itself and carries the end points into themselves; we do not assume that the transformation is 1—1 necessarily. Then we can find a hemicactoid \bar{H}' with base \bar{B}' and cactoids \bar{C}'_m such that (1) there are 1—1 continuous transformations $S_{m,n}^{(1)}$ of $\bar{C}_{m,n}$ into $\bar{C}'_{m,n}$ and $S_n^{(2)}$ of \bar{B}_n into \bar{B}'_n which are conformal (with the obvious conventions at vertices) between corresponding non-degenerate cyclic elements and (2) the transformations $[S_{m,n}^{(1)}]^{-1} T_{m,n}^{(1)}$ and $[S_n^{(2)}]^{-1} T_n^{(2)}$, defined respectively on $\bar{C}'_{m,n}$ and \bar{B}'_n , unite to form a continuous monotone transformation of \bar{H}' into \bar{H} which carries \bar{B}' into \bar{B} .*

Proof. Let $\bar{B}'_1 = \bar{B}_1$. By Lemma 6, we can attach a simple cyclic chain \bar{B}'_2 homeomorphic to \bar{B}_2 at a point of \bar{B}'_1 which is carried by $(S_1^{(2)})^{-1} T_1^{(2)}$ ($S_1^{(2)}$ being taken to be the identity) into the point $\bar{B}_1 \cdot \bar{B}_2$; clearly we may assume that the transformation $S_2^{(2)}$ of \bar{B}_2 into \bar{B}'_2 is conformal between corresponding non-degenerate cyclic elements. It is clear by Lemma 6 that

this can be continued indefinitely for the \bar{B}_n and then for the $\bar{C}_{m,n}$ arranged in a single sequence. By arguments already used, all the conclusions may be demonstrated.

3. *Monotone transformations in the plane.* In this section, we shall merely develop a few important theorems about monotone transformations of one plane set into another.

LEMMA 1. Let $w = f(z)$, $f(z)$ being analytic in R_0 , the unit circle except for a possible single pole, carry R_0 in a 1—1 conformal way into a simply connected region D which may be infinite. Then a necessary and sufficient condition that $f(z)$ be continuous on R_0^* is that D^* be a bounded continuous curve. If this is the case, no continuum of R_0^* corresponds to a point of D^* .

Proof. That the condition is necessary is obvious.

Let us assume, then, that D^* is a bounded continuous curve. Then every point of D^* is accessible from D .† Thus each prime end ‡ of D^* consists of a single point. From the results of Caratheodory,‡ one immediately concludes that $f(z)$ is continuous over R_0^* and that no continuum of R_0^* corresponds to a point of D^* .

LEMMA 2. Let $U = U(u)$ be a monotone transformation carrying a Jordan region \bar{r} , into a base set \bar{B} , in which no continuum over which $U(u)$ is constant separates Π . Then we can find a monotone transformation $U = U'(u)$, of the whole plane Π , into itself such that $U'(u) = U(u)$, $u \in \bar{r}$; and $U = U'(u)$ is 1—1 and continuous for $u \in \Pi - \bar{r}$.

Proof. If \bar{B} consists of a single point, the theorem is immediate, for we may define G_1 to be the upper semicontinuous collection of continua consisting of \bar{r} plus all the points of $\Pi - \bar{r}$. From a theorem of R. L. Moore,§ we can find a monotone transformation $U = U''(u)$ of Π into itself carrying G_1 topologically into the points of Π , and \bar{r} into some point U_0'' . If we define $U'(u) = U''(u) + \bar{U} - U_0''$, ($\bar{U} = U(u)$, $u \in \bar{r}$), $U = U'(u)$ satisfies the requirements.

Now if \bar{B} does not consist of a single point, map $\Pi - \bar{r}$ in a 1—1 conformal way on $\Pi - \bar{r}$ by a function $U = U''(u')$; $U = U''(u')$ is continuous on r^* and no continuum of r^* corresponds to a point of $(\Pi - \bar{B})^*$, by Lemma 1. Now it is well known that there exists a monotone transformation of Π into itself, which is 1—1 and continuous off of r^* , and

† R. L. Wilder, *loc. cit.*, pp. 350-351.

‡ C. Caratheodory, "Die Begrenzung einfache zusammenhängender Gebiete," *Mathematische Annalen*, vol. 73 (1913), pp. 323-351 and footnote, page 251, especially.

§ R. L. Moore, *loc. cit.* (not P. S. T.).

carries any upper semicontinuous collection of continua on r^* , none of which separates r^* , topologically into the points of r^* . Now, the collection of maximal continua of r^* each of which corresponds to a point of $(\Pi - \bar{B})^*$ is upper semicontinuous and no one includes all of r^* or separates r^* . Hence we have such a transformation $u' = u'(u)$, such that $U''(u') = U(u)$, $u \in r^*$. Hence if we define $U'(u) = U''[u'(u)]$, $u \in \Pi - \bar{r}$, and $U'(u) = U(u)$, $u \in \bar{r}$, then $U'(u)$ is continuous and monotone over all of Π and satisfies the other requirements.

Definition 1. By the oscillation of a vector function $\phi(u)$, over a set E , we mean the least upper bound of $|\phi(u_1) - \phi(u_2)|$ for all pairs (u_1, u_2) of points of E .

THEOREM 1. Let $T : U = U(u)$, be a continuous monotone transformation of an open simply connected region r , into an open simply connected region R , and let G be the collection of maximal continua over each of which $U(u)$ is constant. Then

- (i) no continuum of G separates r ;
- (ii) the points of any closed Jordan region \bar{d} , are carried by T into points of the continuum $\bar{D} = \bar{D} + \bar{D}^*$, where $\bar{D}^* = T(d^*)$ and all points (if any) of \bar{D} are separated from ∞ by \bar{D}^* ;
- (iii) if \bar{d} is any closed region in r , the oscillation of $U(u)$ over \bar{d} is equal to its oscillation over d^* ; and
- (iv) all the points of r carried into a simply connected open region D , of R , form a simply connected open region d , of r .

If we replace r and R by closed Jordan regions \bar{r} and \bar{R} , in the hypotheses, all the above conclusions hold, and we may add that no continuum of G contains all of r^* .

Proof. (i) is immediate since T is a topological transformation of G into R and since by Lemma 2, § 2, the set of points of r belonging to a connected set of elements of G is connected.

To demonstrate (ii), let P be a point of R which can be joined to a point Q of R^* by an arc Γ (open at one end) containing no point of \bar{D}^* . To this arc corresponds a connected set γ , of points of r having no point in common with d^* . Let the arc Γ be the outer limit of arcs $\Gamma_1 + \Gamma_2 + \dots$, where $\Gamma_{n+1} \supset \Gamma_n$ and all have P as one end point. Then γ is the outer limit of γ_n , where γ_n is the continuum corresponding to Γ_n . Now γ is not closed since Γ is not; hence if γ were entirely interior to d^* , it would have some limit points in \bar{d} and from the continuity of $U(u)$, these would be carried into Q

which is impossible. Hence $\bar{\gamma}$ lies outside d^* so that all the points corresponding to P lie outside d^* . This proves (ii) since \bar{Q} may be joined to ∞ by a continuum containing no points of R , and (iii) follows immediately from (ii).

Now it is well known that in any plane transformation all the points of the original set corresponding to an open set in the transformed set form a set open with respect to the original set. Thus, to any simply connected open region D , of R corresponds a connected open region d , of r . Now, let γ be a simple closed curve in d ; $\Gamma = T(\gamma)$ lies in D . By (ii) and the simple connectivity of D , it follows that the interior of γ is carried into a subset of D . Thus d is simply connected.

The proof of (i) is the same if r and R are replaced by the Jordan regions \bar{r} and \bar{R} respectively. Let G be replaced by \bar{G} and suppose a continuum \bar{g} , of \bar{G} contains r^* . Now $\bar{r} - \bar{g}$ is obviously open, connected (by i), and is simply connected, since its boundary is a continuum clearly. Thus we may carry \bar{r} in a continuous monotone way into the sphere, the transformation carrying \bar{g} into a point being 1 — 1 otherwise. This determines a collection \bar{G}_1 on the sphere which is topologically equivalent to \bar{G} , and of which no continuum separates the sphere again using (i). From a theorem of R. L. Moore (P. S. T., Theorem 19, chapter V), such a collection is topologically equivalent to the sphere. But \bar{G} was given topologically equivalent to \bar{R} which contradicts the above. Thus no continuum of \bar{G} contains r^* . Having shown this, we may extend \bar{T} to the whole of Π , \bar{T} being 1 — 1 for u outside \bar{r} . By letting r_1 be an open Jordan region, containing \bar{r} and by considering the extended transformation in this region, we see that the remaining results may be easily shown.

THEOREM 2. *Let $T, T : U = U(u), u \in r$, be a continuous monotone transformation of an open simply connected region r into another open simply connected region R . Then there exists a sequence $\{T_n\}$, $T_n : U = U_n(u)$, of 1 — 1 continuous transformations of r into R such that the $U_n(u)$ converge uniformly to $U(u)$ on each closed subset of r . If we replace r and R by Jordan regions \bar{r} and \bar{R} respectively, we may choose the sequence $\{T_n\}$ so that the $U_n(u)$ converge uniformly on the whole of \bar{r} .*

Proof. It is clear that we may assume R to be the interior of Q , the square $(0, 0; 1, 1)$. Define $C_{i,j}$ as the line $V = j/2^i$, $D_{i,j}$ as the line $U = j/2^i$, $c_{i,j}$ as the connected set of r corresponding to $C_{i,j}$, and $d_{i,j}$ the connected set of r corresponding to $D_{i,j}$ ($i = 1, 2, \dots; j = 1, 2, \dots, 2^{i-1}$). Now, for each i , $\Gamma_i = r - \sum_{j=1}^{2^{i-1}} c_{i,j}$ and $\Delta_i = r - \sum_{j=1}^{2^{i-1}} d_{i,j}$ each consist of 2^i

mutually separated simply connected regions $\gamma_{i,j}$ and $\delta_{i,j}$ respectively ($j=1, \dots, 2^i$); let $r_{j,k}^i$ be the simply connected region (corresponding to the open square $R_{j,k}^i$ of R) $\gamma_{i,j} \cdot \delta_{i,j}$. In each $r_{j,k}^i$ ($j, k=1, \dots, 2^i-1$) (i. e., omit the last row and column), choose a point $p_{j,k}^i$, and, for each j , $1 \leq j \leq 2^i-1$, let $\tilde{c}_{i,j}$ be an open curve which passes through all the $p_{j,k}^i$ in $\gamma_{i,j}$, lies entirely interior to $\gamma_{i,j}$, divides it into exactly two parts, but is such that no segment of it divides $\gamma_{i,j}$; and for each k , $1 \leq k \leq 2^i-1$, let $\tilde{d}_{i,k}$ be a similar curve in $\delta_{i,k}$ which has only the point $p_{j,k}^i$ in common with the curve $\tilde{c}_{i,j}$ for each j . Now it is clear that we may find a $1-1$ continuous transformation of r into R which carries $\tilde{c}_{i,j}$ into $C_{i,j}$ and $\tilde{d}_{i,j}$ into $D_{i,j}$. We define $T_i, T_i: U=U_i(u)$, to be any such transformation.

Now let p be any point of r , c_p the continuum of the collection corresponding to $U(u)$ containing it, and P the point of R corresponding to it under T . Let C_1^i, C_2^i, D_1^i , and D_2^i be division lines of the i -th division of $R (=Q)$ such that, for each i , P is interior to the rectangle R_i cut off by the lines and such that $P = \bigcap_{i=1}^{\infty} R_i$. Let c_j^i, d_j^i , and r^i be the sets in r corresponding under T to C_j^i, D_j^i , and R_i , respectively ($j=1, 2$); and let $\tilde{c}_{n,j}^i, \tilde{d}_{n,j}^i$, and $\tilde{r}_{i,n}$ be the corresponding sets under T_n ($n \geq i$), the $\tilde{c}_{n,j}^i$ and $\tilde{d}_{n,j}^i$ being among the $\tilde{c}_{n,j}$ and $\tilde{d}_{n,j}$. Then, for each i , there exists an $N_{i,p}$ such that, for $n > N_{i,p}$, $p \in c_p \subset \tilde{r}_{i,n}$ and hence $P_n = T_n(p) \in R_i$. Hence the sequence $\{U_i(u)\}$ converges for each u . Now, if we restrict ourselves to the region \tilde{r}_n with its boundary which corresponds under T to the closed square $1/2^n \leq u, v \leq 1 - 1/2^n$, we see that the convergence is uniform. For it is clear that we can find a $\delta_i > 0$ such that if $|u_p - u_q| < \delta_i$, p and q ($p, q \in \tilde{r}_n$) are necessarily within some "cluster of four" of the $\tilde{r}_{j,k}^i$ corresponding to four of the $\tilde{R}_{j,k}^i$ in a cluster of the form \mathbb{H} , and hence within a cluster of nine of the $\tilde{r}_{j,k}^{i,m} = T_m^{-1}(\tilde{R}_{j,k}^i)$. Thus for $m > i$,

$$|U_{p_m} - U_{q_m}| < 3(2)^{1/2}/2^i \quad \text{if} \quad |u_p - u_q| < \delta_i.$$

4. *The representation of surfaces on hemicactoids.* In this section, it is first shown that, if \bar{H} is any hemicactoid and $X(U)$ is a continuous vector function defined on \bar{H} , (1) there exists a monotone transformation $U = U(u)$ of a Jordan region \bar{r} into \bar{H} which carries the subcollection G_0 of the corresponding collection G into \bar{B} , and (2) if we define $x(u) = X[U(u)]$, then $x = x(u)$ is a surface and all surfaces thus obtained are identical in the sense of Fréchet. This suggests the possibility of allowing surfaces to be represented on hemicactoids. The remainder of the section then treats of the fundamental questions of such a theory of surfaces.

We have defined the Fréchet distance $\|S_1, S_2\|$ between two surfaces S_1 and S_2 in the introduction. The following lemma is an immediate consequence of these definitions.

LEMMA 1. (a) $\|S_1, S_2\| = \|S_2, S_1\|$; (b) $\|S_1, S_3\| \leq \|S_1, S_2\| + \|S_2, S_3\|$; (c) if $\|S_1, \bar{S}_1\| = \|S_2, \bar{S}_2\| = 0$, then $\|S_1, S_2\| = \|\bar{S}_1, \bar{S}_2\|$; and (d) if $\lim_{n \rightarrow \infty} \|S, S_n\| = 0$ and $x = x(u)$ is any representation of S , we can find representations $x = x_n(u)$, of S_n such that the $x_n(u)$ converge uniformly to $x(u)$.

Definition 1. If $\lim_{n \rightarrow \infty} \|S, S_n\| = 0$, we say that S_n approaches S or $\lim_{n \rightarrow \infty} S_n = S$.

LEMMA 2. Let $T, T : U = U(u)$, and $T_1, T_1 : U = U_1(u)$, be two continuous monotone transformations of \bar{r} into a base set \bar{B} , and G and G_1 the collections of maximal continua over each of which $U(u)$ and $U_1(u)$, respectively, is constant (hereafter, we shall call these collections, the collections corresponding to the given monotone transformations). If no continuum of either G or G_1 separates the plane and we define the "surfaces" S and S_1 by $S : U = U(u)$, $S_1 : U = U_1(u)$, then S and S_1 are identical in the sense of Fréchet.

Proof. Let T and T_1 be extended monotonely to the whole of Π , each being 1 — 1 outside of \bar{r} . Let \bar{R}_n be a sequence of closed Jordan regions such that $\bar{R}_n \supset \bar{R}_{n+1}$ for each n and $\prod_{n=1}^{\infty} \bar{R}_n = \bar{B}$; and let \bar{r}_n be $T^{-1}(\bar{R}_n)$ and \bar{r}_{1n} be $T_1^{-1}(\bar{R}_n)$. It is clear that the Jordan closed curves r_n^* and r_{1n}^* approach r^* uniformly and hence also that $S_n \rightarrow S$ and $S_{1n} \rightarrow S_1$ respectively where S_n and S_{1n} are defined by

$$S_n : U = U(u), u \in \bar{r}_n, \quad S_{1n} : U = U_1(u), u \in \bar{r}_{1n}.$$

But now, by Theorem 2, § 3, there exists a sequence $\{T_m\}$, $T_m : U = U_m(u)$, of 1 — 1 continuous transformations of \bar{r}_n into \bar{R}_m which approaches T uniformly. Thus it follows immediately (from the definition of a surface) that $S_n = S'_n$, where $S'_n : U' = U, U \in \bar{R}_n$. Similarly $S_{1n} = S'_n$, so that $S_n = S_{1n}$, for each n . Thus it is easy to see that $S = S_1$.

LEMMA 3. Let $S, S : x = x(u)$, $u \in \bar{r}$, be a surface and let \bar{g} be a continuum of \bar{r} which corresponds to a point of S and possesses a finite number of bounded (simply connected) complementary domains D_1, \dots, D_n . Let C_1 be a simple closed curve in D_1 and R_1 the Jordan subregion of D_1 bounded by C_1 . Let T_1 be a 1 — 1 continuous transformation of D_1 into R_1 and let

S_1 be represented by $x = x_1(\bar{u})$, $\bar{u} \in \bar{r}$, where $x_1(\bar{u}) = x(u)$, u and \bar{u} being identical if u is outside \bar{D}_1 , u and \bar{u} corresponding under T_1 if $u \in D_1$, or $\bar{u} \in \bar{D}_1 - R_1$ and $u \in D^*_{11}$, simultaneously.

Then $x_1(\bar{u})$ is continuous and S_1 is a surface identical with S .

Proof. It is easily seen that $x_1(\bar{u})$ is continuous.

Let $R_1^{(m)}$ be a sequence of open Jordan regions where $R_1^{(m)} \subset R_1^{(m+1)}$ for each m and $D_1 = \sum_{m=1}^{\infty} R_1^{(m)}$. Let $T_1^{(m)}$ be a 1—1 continuous transformation of \bar{R}_1 into $\bar{R}_1^{(m)}$ and suppose that the sequence $\{T_1^{(m)}\}$ converges uniformly to T_1^{-1} on every closed subregion of R_1 . It is clear that we may define for each m , a 1—1 continuous transformation of $\bar{D}_1 - R_1^{(m)}$ into $\bar{D}_1 - \bar{R}_1$ which is the identity near D^*_{11} and which preserves a preassigned correspondence between R^*_{11} and $R^{*1(m)}$. Hence we may define a sequence $\{T_m\}$, $T_m: \bar{u} = \bar{u}_m(u)$, of 1—1 continuous transformations of \bar{r} into itself which is the identity for u not in \bar{D}_1 , is $(T_1^{(m)})^{-1}$ for $u \in \bar{R}_1^{(m)}$, and is the correspondence in the preceding sentence for $u \in \bar{D}_1 - R_1^{(m)}$. If we define $x_m(u) = x_1[\bar{u}_m(u)]$, it is easy to see that $\{x_m(u)\}$ converges uniformly to $x(u)$. But each of the surfaces S_m , $S_m: x = x_m(u)$, is identical with S_1 and by the preceding sentence, $\lim_{m \rightarrow \infty} \|S, S_m\| = 0$. Hence $S = S_1$.

LEMMA 4. Let \bar{C} be a double cone and $X(U)$ a continuous vector function defined on it. Suppose there exist two monotone transformations T , $T: U = U(u)$, and T_1 , $T_1: U = U_1(u)$, of \bar{r} into \bar{C} , r^* being carried by each into the same endpoint of \bar{C} . Let S and S_1 be defined by

$$S: x = x(u) = X[U(u)], \quad S_1: x = x_1(u) = X[U_1(u)].$$

Then $S = S_1$.

Proof. Let G and G_1 be the collections of continua in \bar{r} corresponding to T and T_1 respectively and let g^* and g^*_{11} be the continua of G and G_1 , respectively, which contain r^* . Now, clearly (using Lemma 2, § 2 and the fact that g^* and g^*_{11} are continua), the regions $D = \bar{r} - g^*$ and $D_1 = \bar{r} - g^*_{11}$ are simply connected. Now, let $U = U_2(v)$ be a monotone transformation of the unit circle \bar{R}_0 into \bar{C} which carries R^*_0 into the above endpoint and is otherwise 1—1 and let S_2 be represented on \bar{R}_0 by $x = \chi(v) = X[U_2(v)]$. Clearly there are "induced" monotone continuous transformations $v = v(u)$, and $v = v_1(u)$, of D and D_1 , respectively, into $R_0 (= \bar{R}_0 - R^*_0)$ such that $U(u) = U_2[v(v)]$ and $U_1(u) = U_2[v_1(u)]$ for u in D and D_1 respectively. Now there exist sequences $\{\tau_n\}$ and $\{\tau_{1,n}\}$, $\tau_n: v = v_n(u)$, $\tau_{1,n}: v = v_{1,n}(u)$, of 1—1 continuous transformations of D and D_1 respectively, into R_0 , such

that $\{v_n(u)\}$ and $\{v_{1,n}(u)\}$ converge uniformly to $v(u)$ and $v_1(u)$ on all closed subsets of D and D_1 . Now let $\{\bar{D}_m\}$ and $\{\bar{D}_{1m}\}$ be sequences of closed Jordan regions such that $\bar{D}_m \subset \bar{D}_{m+1}$ and $\bar{D}_{1m} \subset \bar{D}_{1,m+1}$ for each m , and $D = \sum_{m=1}^{\infty} \bar{D}_m$, $D_1 = \sum_{m=1}^{\infty} \bar{D}_{1m}$. Now, let S_m and S_{1m} be defined by

$$S_m : x = x(u), \quad u \in \bar{D}_m, \quad S_{1m} : x = x_1(u), \quad u \in \bar{D}_{1m}.$$

Clearly $\{S_m\}$ and $\{S_{1m}\}$ converge to S and S_1 , respectively. Now, for each m , let n_m be so large that $|v_{n_m}(u) - v(u)| < 1/m$, $u \in \bar{D}_m$; and

$$|v_{1,n_m}(u) - v_1(u)| < 1/m, \quad u \in \bar{D}_{1m},$$

and let \bar{R}_m and \bar{R}_{1m} be the Jordan regions in R_0 corresponding, under τ_{n_m} to \bar{D}_m and \bar{D}_{1m} , respectively. Let S'_m and S'_{1m} be defined by

$$\begin{aligned} S'_m : x &= x_m(u) = X\{U_2[v_{n_m}(u)]\}, \quad u \in \bar{D}_m, \\ S'_{1m} : x &= x_{1m}(u) = X\{U_2[v_{1,n_m}(u)]\}, \quad u \in \bar{D}_{1m}. \end{aligned}$$

Since $x(u) = X\{U_2[v(u)]\}$ and $x_1(u) = X\{U_2[v_1(u)]\}$, it is easy to see that S'_m and S_m approach the same limit. Since $S'_m = S''_m$, $S''_m : x = X[U_2(v)]$, $v \in \bar{R}_m$, and since it is clear that \bar{R}_m converges to R^*_0 , we see that $\{S'_m\}$ converges to S_2 . Similarly, $\{S_{1m}\}$ and $\{S'_{1m}\}$ converge to the same limit, $\{S'_{1m}\}$ converging to S_2 . Thus $S = S_2 = S_1$, which proves the lemma.

LEMMA 5. *Let the hypotheses of Lemma 4 be satisfied with \bar{C} being any simple cyclic chain of type A instead of merely a double cone. Then also, $S_1 = S$.*

Proof. First of all, let P and Q be two cut points of any simple cyclic chain \bar{C}' , of type A and suppose P separates Q from an endpoint E , of \bar{C}' . Let T' be a continuous monotone transformation of \bar{r} into \bar{C}' in which r^* is carried into E , and let \bar{G}' be the corresponding collection of continua in \bar{r} . Now, using Lemma 2, § 2, we see that (1) the continua g'_P and g'_Q of G' corresponding to P and Q respectively, each bound exactly one simply connected region, say d'_P and d'_Q , (2) $g'_Q \subset d'_P$, (3) the region $r'_{P,Q} = d'_P - g'_Q - d'_Q$, between g'_P and g'_Q , is doubly connected, and (4) $g'_P + r'_{P,Q} + g'_Q$ can be mapped in a monotone continuous way on to the surface of a sphere, g'_P and g'_Q being carried into any two distinct points, the mapping being 1—1 and continuous otherwise. Hence, if we replace the part of \bar{C}' between P and Q by another simple cyclic chain $\bar{C}''_{P,Q}$ of type A with endpoints P and Q , we can find an upper semicontinuous collection, $G''_{P,Q}$ filling up $r'_{P,Q}$ so that $\bar{G}''_{P,Q} = G''_{P,Q} + g'_P + g'_Q$ and $\bar{G}' - G'_{P,Q} + G''_{P,Q}$ are upper semicontinuous collections topologically equivalent to $\bar{C}''_{P,Q}$ and $\bar{C}' - \bar{C}'_{P,Q} + \bar{C}''_{P,Q}$ respec-

tively, where $C'_{P,Q}$ is the part of \bar{C}' between P and Q , $\bar{C}'_{P,Q}$ and $G'_{P,Q}$ is the subset of G' filling up $r'_{P,Q}$ and correspond can be done in virtue of (4) above and the fact that a coll found on the sphere topologically equivalent to $\bar{C}''_{P,Q}$ is correspond to points on the sphere (using Lemma 9, § 2).

Hence, form \bar{C}_m from \bar{C} by replacing each double co $< 1/m$ by its axis, and define $X_m(U)$ on \bar{C}_m by letting $U \in \bar{C} \cdot \bar{C}_m$, and $X_m(U) = X(U')$, for U on the axis of o double cones of \bar{C} , U' ranging linearly over the broken lin two equal sides of a generating triangle of that double con the axis. In each region in \bar{r} corresponding to such a doub part of \bar{G} and \bar{G}_1 corresponding to the double cone by a n logically equivalent to the axis. It is easily seen that the n upper semicontinuous. Hence let $U = U_m(u)$ and $U = U$ sponding monotone transformations of \bar{r} into \bar{C}_m ; and l defined by

$$S_m : x = x_m(u) = X_m[U_m(u)], \quad S_{1m} : x = x_{1m}(u) =$$

Then it is clear that $\lim_{m \rightarrow \infty} S_m = S$ and $\lim_{m \rightarrow \infty} S_{1m} = S_1$, notic and $\{U_{1m}(u)\}$ converge uniformly to $U(u)$ and $U_1(u)$, , let $\bar{C}_{m,n}$ be formed from \bar{C}_m by dividing each interval of \bar{C}_m double cones into n equal parts and replacing each subinterv with the same end points and the proper vertex angle. I $\bar{C}_{m,n}$ by the relations

$$X_{m,n}(U) = X_m(U), U \in \bar{C}_m \cdot \bar{C}_{m,n}, X_{m,n}(U) = X_m(U'), U$$

U' being the intersection, with the axis of $\bar{C}_{m,n}$ of a pla perpendicular to the axis. Now let $U = U_{m,n}(u)$ and monotone transformations of \bar{r} into $\bar{C}_{m,n}$ coinciding with U except on regions corresponding to $\bar{C}_m - \bar{C}_m \cdot \bar{C}_{m,n}$ and def as indicated above. Now, again, $\lim_{n \rightarrow \infty} S_{m,n} = S_m$ and $\lim_{n \rightarrow \infty} S_{1n}$

we shall prove our theorem for chains \bar{C} made up entirely of abutting double cones from which the general case will processes.

We shall prove the theorem for this case by induction of double cones. We have proved the theorem when n . Suppose the theorem is true for $n = k$ and that \bar{C} has k Let \bar{C}_k consist of the first k double cones of \bar{C} , starting fro other end point of \bar{C}_k , and let g_k and g_{1k} be the continua

to E_k under T and T_1 , respectively and d_k and d_{1k} the regions bounded, respectively, by g_k and g_{1k} . Let C_k be a simple closed curve in d_k and C_{1k} one in d_{1k} and let R_k be the Jordan subregion of d_k bounded by C_k and R_{1k} that of d_{1k} bounded by C_{1k} . Let T_k be a 1—1 continuous transformation of d_k into R_k and T_{1k} one of d_{1k} into R_{1k} . Let $\bar{x}(\bar{u})$ and $\bar{x}_1(\bar{u}_1)$ be defined for \bar{u} and \bar{u}_1 on \bar{r} by the relations $\bar{x}(\bar{u}) = x(u)$ and $\bar{x}_1(\bar{u}_1) = x_1(u)$, where $u = \bar{u}$, $u \in \bar{r} - \bar{d}_k$, u and \bar{u} correspond under T_k if $u \in d_k$, or $u \in d_k^*$ and $\bar{u} \in \bar{d}_k - R_k$, simultaneously, an analogous statement holding for u and \bar{u}_1 . Then $x = \bar{x}(\bar{u})$ and $x = \bar{x}_1(\bar{u}_1)$, $\bar{u}, \bar{u}_1 \in \bar{r}$, are representations of S and S_1 , respectively, by Lemma 3. Now, let S_k and S_{1k} be given respectively by $x = x_k(u)$ and $x = x_{1k}(u)$, where

$$x_k(u) = x(u), u \in \bar{r} - d_k, \quad x_k(u) = X(U_{E_k}), u \in d_k,$$

and $x_{1k}(u)$ is defined analogously. By Lemma 3, S_k and S_{1k} are also given by $x = \bar{x}_k(\bar{u})$ and $x = \bar{x}_{1k}(\bar{u}_1)$, respectively, where $\bar{x}_k(\bar{u})$ and $\bar{x}_{1k}(\bar{u}_1)$ are obtained from $\bar{x}(\bar{u})$ and $\bar{x}_1(\bar{u}_1)$ in the way that $x_k(u)$ and $x_{1k}(u)$ are obtained from $x(u)$ and $x_1(u)$, $\bar{r} - d_k$ and $\bar{r} - d_{1k}$ being replaced by $\bar{r} - R_k$ and $\bar{r} - R_{1k}$, respectively. By hypothesis of the induction $S_k = S_{1k}$, so that there exists a sequence $\{\bar{\tau}_{k,n}\}$ of 1—1 continuous transformations of \bar{r} into itself, and hence a sequence $\{\tau_{k,n}\}$, $\tau_{k,n} : \bar{u}_1 = \bar{u}_1^{(n)}(\bar{u})$, of such transformations of $\bar{r} - R_k$ into $\bar{r} - R_{1k}$ conserving a given correspondence between C_k and C_{1k} and such that the sequence $\{\bar{x}_n(\bar{u})\}$, defined on $\bar{r} - R_k$ by $\bar{x}_n(\bar{u}) = \bar{x}_{1k}[\bar{u}_{1n}(\bar{u})]$, converges uniformly on $\bar{r} - R_k$ to $\bar{x}(\bar{u})$. Now let σ_k and σ_{1k} be the surfaces defined on \bar{r} by $x = x'_k(u)$ and $x = x'_{1k}(u)$, respectively, where

$$x'_k(u) = x(u), u \in d_k, \quad x'_k(u) = X(U_{E_k}), u \in \bar{r} - d_k,$$

and $x'_{1k}(u)$ is defined analogously. By Lemma 3, σ_k and σ_{1k} are represented on \bar{R}_k and \bar{R}_{1k} , respectively, by $x = \bar{x}(\bar{u})$, $\bar{u} \in \bar{R}_k$, and $x = \bar{x}_1(\bar{u}_1)$, $\bar{u}_1 \in \bar{R}_{1k}$. By Lemma 4, $\sigma_k = \sigma_{1k}$ (the part of S and S_1 corresponding to the last double cone of \bar{C}) so there exists a sequence $\{\tau'_{k,n}\}$, $\bar{u}_1 = \bar{u}'_1{}^{(n)}(\bar{u})$, of 1—1 continuous transformations of \bar{R}_k into \bar{R}_{1k} preserving the above correspondence between R_k^* and R_{1k}^* , such that the sequence $\{\bar{x}_n(\bar{u})\}$, defined on \bar{R}_k by $\bar{x}_n(\bar{u}) = \bar{x}_1[\bar{u}'_1{}^{(n)}(\bar{u})]$, converges uniformly on \bar{R}_k to $\bar{x}(\bar{u})$. Now, for each n , $\tau_{k,n}$ and $\tau'_{k,n}$ unite to form a 1—1 continuous transformation $T'_{k,n}$, $T'_{k,n} : \bar{u}_1 = \bar{u}_1^{(n)}(\bar{u})$, of \bar{r} into itself and the above $\bar{x}_n(\bar{u})$ is seen to be continuous over \bar{r} and also given by $\bar{x}_n(\bar{u}) = \bar{x}_1[\bar{u}_1^{(n)}(\bar{u})]$. Hence $S = S_1$ and our induction is complete.

THEOREM 1. *Let \bar{H} be a hemicactoid and $X(U)$ a continuous vector function defined on it. There exists a continuous monotone transformation $U = U(u)$, with associated upper semicontinuous collection G , of a Jordan*

region \bar{r} , into \bar{H} which carries the subcollection G_0 , of \bar{H} logically into \bar{B} . If we define $x(u) = X[U(u)]$, then x and all surfaces obtained in this way are identical.

Proof. The first part of the theorem has been proved.

To demonstrate the last part, let $T, T: U = U(u)$, and $u \in \bar{r}$, be two continuous monotone transformations with collections G and G_1 , which carry the above subcollections. Let S and S_1 be the corresponding surfaces obtained as a result of \bar{H} by

$$\bar{H} = \bar{B} + \sum_{n=1}^{\infty} \bar{C}_n + H^*,$$

where, for each n , \bar{C}_n is a simple cyclic chain of $\bar{C}_n \cdot [\bar{B} + \sum_{j=1}^{n-1} \bar{C}_j]$ is an end point of \bar{C}_n , and H^* is a compact set of limit points of $\bar{B} + \sum_{j=1}^{\infty} \bar{C}_j$. Let $\bar{H}_0 = \bar{B}$, $\bar{H}_n = \bar{B} + \sum_{j=1}^n \bar{C}_j$ be the subcollections of G and G_1 , respectively, corresponding to T_0 and T_1 to \bar{H}_n , and let \hat{G}_n and \hat{G}_{1n} denote the respective regions covered by these collections. Now it is easy to see (using Lemma 8, § 2) that \hat{G}_n is a continuum, that every point on \hat{G}_n or is in a bounded complementary domain of a collection \hat{G}_n that the same statements hold for \hat{G}_{1n} . Thus, for each n , let g'_n be a collection of continua g'_n , each g'_n being obtained from \hat{G}_n by all the complementary domains of \hat{G}_n which it bounds, and g'_{1n} similarly formed from \hat{G}_{1n} . Let $T_n, T_n: U = U_n(u)$, and T_{1n} , be the monotone continuous transformations of \bar{r} into \bar{H}_n , G'_n and G'_{1n} as their respective corresponding collections. Let S_n and

$$S_n: x = x_n(u) = X[U_n(u)], \quad S_{1n}: x = x_{1n}(u)$$

Then clearly the sequences $\{U_n(u)\}$ and $\{U_{1n}(u)\}$, converge to $U(u)$ and $U_1(u)$ respectively, and $\lim_{n \rightarrow \infty} S_n = S$ and $\lim_{n \rightarrow \infty} S_{1n} = S_1$ if we prove the theorem for the case when the number of continua is finite. The general case will follow by a simple limit process. But the case may be handled by induction in precisely the same manner as the corresponding proof in Lemma 5 was carried through.

The above theorem gives a logical justification for the statement "on hemicactoids". The remainder of this section is devoted to certain essential parts of a theory of such representations:

Definition 2. The two surfaces $x = X(U)$, $U \in \bar{H}$,

\bar{H} being a hemicactoid, will be said to be *identical* if there exists a continuous monotone transformation $U = U(v)$, $v \in \bar{K}$, of a Jordan region \bar{K} into \bar{H} , with corresponding collection G , carrying G_0 topologically into \bar{B} , and such that the surface $x = \xi(v) = X[U(v)]$, $v \in \bar{K}$, is identical with the surface $x = x(u)$, $u \in \bar{\tau}$. By the above theorem, all the surfaces $x = \xi(v)$ thus obtained are identical.

Definition 3. The vector function $X(U)$, $U \in \bar{H}$, \bar{H} a continuum, is *non-degenerate* on \bar{H} if $X(U)$ is not constant over any continuum of \bar{H} containing two points.

Definition 4. The representation $x = X(U)$, $U \in \bar{H}$, of a surface on the hemicactoid \bar{H} , is *non-degenerate* if $X(U)$ is non-degenerate on \bar{H} .

THEOREM 2. Any surface S , $S : x = x(u)$, $u \in \bar{\tau}$, can be represented non-degenerately on some hemicactoid. If $x = X(U)$, $U \in \bar{H}$, and $x = X_1(U_1)$, $U_1 \in \bar{H}_1$, are two such representations of S , then there exists a 1—1 continuous transformation $U_1 = U_1(U)$, of \bar{H} into \bar{H}_1 carrying \bar{B} into \bar{B}_1 such that $X_1[U_1(U)] = X(U)$.

Proof. Let G be the upper semicontinuous collection of maximal continua over each of which $x(u)$ is constant. Now, by Theorem 3, § 2, we may carry G topologically into a hemicactoid \bar{H} , by a monotone continuous transformation $U = U(u)$, which also carries G_0 into \bar{B} , and if $U_1 = U_1(u)$ is another such transformation, carrying G topologically into a hemicactoid \bar{H}_1 and G_0 into \bar{B}_1 then there exists a 1—1 continuous transformation $U_1 = \bar{U}_1(U)$, of \bar{H} into \bar{H}_1 and \bar{B} into \bar{B}_1 such that $\bar{U}_1[U(u)] = U_1(u)$. Hence if we define $X(U)$ on \bar{H} so that $X[U(u)] = x(u)$ as can obviously be done, we have only to show that $X(U)$ is continuous and non-degenerate in order to prove the theorem. But this is immediate, for, let $\{U_n\}$, $U_n \in \bar{H}$, converge to U , and let $\{g_n\}$ and g be the corresponding continua in $\bar{\tau}$. Then by Theorem 2, § 2, g is the limit continuum of $\{g_n\}$ and there exists a sequence of points $\{p_n\}$, $p_n \in g_n$, such that every convergent subsequence converges to a point p of g ; this proves the continuity of $X(U)$. The non-degeneracy is obvious.

Definition 5. A point P , of a hemicactoid \bar{H} , is said to be separated from the point ∞ in the plane Π , of \bar{B} by a set σ if every connected set in \bar{H} containing P and a point of $(\Pi - \bar{B})^*$ also contains a point of σ .

THEOREM 3. If $x = X(U)$, $U \in \bar{H}$, is a non-degenerate representation of S , and $x = x(u)$, $u \in \bar{h}$, is any other representation of S (\bar{H} and \bar{h} hemicactoids) there exists a continuous monotone transformation $U = U(u)$, with corresponding collection of continua $G^{(\bar{h})}$, which carries \bar{h} into \bar{H} , is such that

$X[U(u)] = x(u)$, and carries the collection $G_0^{(\bar{h})}$, topologically into \bar{B} , $G_0^{(\bar{h})}$ consisting of the continua of $G^{(\bar{h})}$, no point of any of which is separated from the point ∞ in the plane of \bar{b} by a continuum of $G^{(\bar{h})}$. If $x = x(u)$ is also non-degenerate, any such transformation is 1 — 1. The collection $G^{(\bar{h})}$ may be taken to be the collection of maximal continua of \bar{h} over each of which $x(u)$ is constant.

Proof. The last two statements are consequences of Theorem 2.

Now take $G^{(\bar{h})}$ as the collection of maximal continua of \bar{h} over each of which $x(u)$ is constant and let $u = u(v)$ be a continuous monotone transformation with corresponding collection Γ , which carries \bar{r} into \bar{h} and Γ_0 topologically into \bar{b} . Now it is easy to verify the fact that $G^{(\bar{h})}$ is carried topologically into the collection G , of maximal continua over each of which $\xi(v) = x[u(v)]$ is constant. Now it is clear that $u = u(v)$ sets up a topological correspondence between the points of $(\Pi - \bar{B})^*$ and those continua of Γ which have points in common with r^* , that each continuum of G corresponds to a continuum of elements of Γ , that $u = u(v)$ carries a connected point set of \bar{r} into one on \bar{h} , and that, to a connected set of points of \bar{h} corresponds one such in \bar{r} . Hence it follows easily that $u = u(v)$ establishes a topological relation between the continua of $G_0^{(\bar{h})}$ and G_0 . Now, let $U_1 = U_1(v)$ be a continuous monotone transformation with corresponding collection G , which carries \bar{r} into \bar{H}_1 and G_0 topologically into \bar{B}_1 , and let $X_1(U_1)$ be determined (see Theorem 2) so $X_1[U_1(v)] = \xi(v)$. Then, by Definition 2, S is given by $x = \xi(v)$, $v \in \bar{r}$, $x = X_1(U_1)$, $U_1 \in \bar{H}_1$, and by hypothesis also by $x = X(U)$, $U \in \bar{H}$. Hence there exists a 1 — 1 continuous transformation $U = U(U_1)$, of \bar{H}_1 into \bar{H} and \bar{B}_1 into \bar{B} such that $X[U(U_1)] = X_1(U_1)$. By these successive transformations we have set up a topological relation $U = U(u)$, of the continua of $G^{(\bar{h})}$ into the points of \bar{H} which is easily seen to possess the required properties.

Definition 6. The transformations $U = U(u)$, obtained this way will be said to be *induced* by the change of representation.

5. *On non-degenerate surfaces.* In this section, we consider a very important class of surfaces, which from some points of view are the most general surfaces to which one would like to apply the name of "surface." A characterization of such surfaces in terms of their given representation is developed and an interesting subclass of such surfaces is pointed out.

Definition 1. A surface is said to be *non-degenerate* if it possesses a non-degenerate representation on a Jordan region.

THEOREM 1. If $x(u)$ is non-degenerate on a Jordan region \bar{r} , and $u = u(U)$ is a 1 — 1 continuous transformation of a Jordan region \bar{R} into \bar{r} and we define $X(U) = x[u(U)]$, then $X(U)$ is non-degenerate on \bar{R} .

Remark. In the case of a non-degenerate surface, it is clear that any hemicactoid on which the surface may be represented non-degenerately consists only of its base set and this of a single rhombus.

THEOREM 2. If (i) $S, S : x = X(U)$, and $S_n, S_n : x = X_n(U)$, $U \in \bar{R}$, are all non-degenerate surfaces, (ii) $\lim_{n \rightarrow \infty} \|S, S_n\| = 0$, and (iii) $x = x(u)$, $u \in \bar{r}$, is a non-degenerate representation of S , we can find a sequence of non-degenerate representations $x = x_n(u)$, of S_n so that the $x_n(u)$ approach $x(u)$ uniformly.

Proof. Each S_n possesses a non-degenerate representation $x = \tilde{x}_n(\bar{U})$, $\bar{U} \in \bar{\bar{R}}$, on some Jordan region $\bar{\bar{R}}$. From the definition of Fréchet distance, it follows that, for each n , we can find a 1 — 1 continuous transformation $\bar{U} = \bar{U}_n(u)$, of \bar{r} into $\bar{\bar{R}}$ such that

$$|x(u) - x_n(u)| < 2 \|S, S_n\|, \quad x_n(u) = \tilde{x}_n[\bar{U}_n(u)].$$

From this and Theorem 1, it is clear that $x = x_n(u)$ is the desired sequence.

THEOREM 3. A necessary and sufficient condition that the surface $S, S : x = x(u)$, $u \in \bar{r}$, be non-degenerate is that (i) no continuum over which $x(u)$ is constant separates \bar{r} and (ii) $x(u)$ is not constant over r^* .

Proof. If S is non-degenerate, there exists a monotone continuous transformation $U = U(u)$, of \bar{r} into the square \bar{Q} which carries the collection G , of maximal continua of \bar{r} over each of which $x(u)$ is constant topologically into the points of \bar{Q} (using Theorem 3, § 4). Now by Theorem 1, § 3, no continuum of G separates \bar{r} or contains all of r^* .

Now suppose these conditions are satisfied by the continua of G . Then we can make a monotone continuous transformation $u_1 = u_1(u)$, of \bar{r} into \bar{r}_1 which is 1 — 1 inside r and carries the intervals (may reduce to points) of r^* over each of which $x(u)$ is constant into points of $r^*_{\bar{1}}$. The collection G , is carried topologically into a collection G_1 , no continuum of which has more than one point in common with $r^*_{\bar{1}}$, since $x_1(u_1) = x(u)$ is not constant over any interval of $r^*_{\bar{1}}$ and if a g_1 contained two points of $r^*_{\bar{1}}$ it would separate \bar{r}_1 and the corresponding g would separate \bar{r} . It is clear that $x = x_1(u_1)$, $u_1 \in \bar{r}_1$, is a representation of S . Let G'_1 be the upper semicontinuous collection consisting of G_1 plus all the points of $\Pi - \bar{r}_1$. There exists † a continuous mono-

† R. L. Moore, *loc. cit.* (not P. S. T.).

tone transformation $U = U_1(u_1)$, which carries the continua of G'_1 topologically into the points of Π . It is clear that r^{*}_1 is carried into a Jordan curve, which we may take to be Q^* , by this transformation. By Theorem 1, § 3, every point inside r_1 is carried inside or on Q^* and every point of r^{*}_1 we know is carried into Q^* . Thus $U = U_1(u_1)$ carries \bar{r}_1 into \bar{Q} for if a point of Q were missed it would be necessary for points of r^{*}_1 to be carried into the interior of Q . Thus $U = U(u) = U_1[u_1(u)]$ carries \bar{r} in a monotone way into \bar{Q} , carrying the continua of G topologically into the points of \bar{Q} . By arguments already used, we see that we can define $X(U)$ on \bar{Q} so as to be continuous and non-degenerate, and so that $X[U(u)] = x(u)$. Since $U = U(u)$ is the uniform limit of a sequence of 1 — 1 continuous transformations of \bar{r} into \bar{Q} , it is easily seen that $x = X(u)$, $U \in \bar{Q}$ is a non-degenerate representation of S .

Definition 2. We say that the surface S , $S : x = x(u)$, $u \in \bar{r}$, is *bounded by a Jordan curve C* , if $x = x(p)$, $p \in r^*$, is a representation of C on r^* , i. e., if it is possible to find a continuous monotone transformation $P = T(p)$, of r^* into another simple closed curve R^* , such that there exists a continuous function $X(P)$, defined on R^* such that (i) $X[T(p)] = x(p)$ and (ii) the equation $x = X(P)$ gives a 1 — 1 continuous representation of C on R^* .

Remark. It is easy to see that in this case, the base set of a hemicactoid on which S is represented non-degenerately, consists of a single rhombus.

Definition 3. A function $f(x, y)$, defined on a region \bar{r} , is said to be *monotone in the sense of Lebesgue* if, for every subregion \bar{d} , of \bar{r} , the oscillation of $f(x, y)$ on \bar{d} is equal to its oscillation on \bar{d}^* .

Definition 4. The surface S , $S : x = x(u)$, $u \in \bar{r}$, is said to be *Lebesgue monotone* if each of the component functions, $x^i(u)$, is monotone in the sense of Lebesgue. It is easily seen that this property of a surface is invariant under changes of representation on Jordan regions.

THEOREM 4. Any hemicactoid on which a Lebesgue monotone surface may be represented non-degenerately consists only of its base set.

Proof. This is immediate, for no continuum over which $x(u)$ is constant can separate the plane so that the collection G of maximal continua of \bar{r} over each of which $x(u)$ is constant is equal to its subcollection G_0 .

THEOREM 5. A monotone surface bounded by a Jordan curve C , is non-degenerate.

Proof. This follows immediately from Theorem 4 and the remark after Definition 2.

CONCERNING SPHERICAL SPACES.†

By LEONARD M. BLUMENTHAL.‡

1. *Introduction.* The n -dimensional spherical space $S_{n,r}$ consists of the points of the surface of an $(n+1)$ -dimensional sphere of radius r in a euclidean space of $n+1$ dimensions, with the distance between two points defined as geodesic (great circle) distance. A recent paper characterized those semi-metric spaces that are congruent (i. e., isometric) with a subset of the $S_{n,r}$.§ If p_1, p_2, \dots, p_n are n points of a semi-metric space and we form the symmetric determinant $\Delta(p_1, p_2, \dots, p_n) = |\cos p_i p_j / r|$, $(i, j = 1, 2, \dots, n)$, Theorems I_n, II_n, III_n of the paper referred to may be combined to give the following theorem.

THEOREM.¶ *A semi-metric space is congruent with a subset of $S_{n,r}$ if and only if, no pair of points of the space has a distance exceeding πr and,*

(i) *for every integer $k \leq n+1$ and every k points of the space, the determinant $\Delta(p_1, p_2, \dots, p_k)$ is positive or zero,*

(ii) *the determinant Δ formed for every $n+2$ points of the space vanishes,*

(iii) *the determinant of every $n+3$ points vanishes.*

This theorem gives, essentially, the conditions that $n+3$ points of a semi-metric space be isometric with $n+3$ points of the $S_{n,r}$, and no assumptions of a qualitative nature are made concerning the semi-metric space containing the $n+3$ points.

The object of this paper is to show that if certain simple qualitative conditions are placed upon the semi-metric space, the chain of determinant conditions contained in hypothesis (i) of the theorem are implied by conditions involving only sets of *four points* of the space, while hypothesis (iii) is superfluous, being a consequence of hypothesis (ii). A semi-metric space satisfying our postulates is shown to be congruent with $S_{n,r}$ or with the spherical Hilbert space of radius r according as the dimension of the space is finite or infinite.

It is to be remarked that the method used in proving the fundamental

† Presented to the American Mathematical Society, March 30, 1934.

‡ National Research Fellow.

§ Blumenthal and Garrett, "Characterization of spherical and pseudo-spherical sets of points," *American Journal of Mathematics*, vol. 55 (1933), pp. 619-640.

¶ This result has also been obtained by L. Klanfer, of Vienna, a brief account of whose work appears in the *Ergebnisse eines Mathematischen Kolloquiums*, Wien, Heft 4 (1933), p. 43.

Theorem 3.3 makes no use of the determinant conditions that express the necessary and sufficient conditions that $n + 1$ points of a semi-metric space be congruent with $n + 1$ points of $S_{n,r}$. Hence the procedure is applicable to those spaces for which analogous determinant conditions are unknown. *The corresponding theorem is valid for a wide class of n -dimensional spaces admitting a rotation about an $(n - 2)$ -dimensional subset.*†

2. A set of undefined elements, for suggestiveness called points, and a positive number r form an S_r space provided that to each pair p_1, p_2 of elements there corresponds a real number p_1p_2 , called their distance, satisfying the following postulates:

Postulate A. $p_1p_2 = 0$, if and only if $p_1 = p_2$.

Postulate B. Every three elements p_1, p_2, p_3 of the space satisfy the triangle inequality; i. e., $p_1p_2 + p_1p_3 \geq p_2p_3$.

Postulate C. For every three elements p_1, p_2, p_3 of the space, the sum of the three distances they determine does not exceed $2\pi r$.

Postulate D. If p_1, p_2, p_3, p_4 are four elements of the space such that some triple contained in the four points is isometric with three points of a circle $S_{1,r}$, of radius r , then the four points are isometric with four points of a two-dimensional spherical surface $S_{2,r}$, of radius r .‡

From Postulates A, B it follows that $p_1p_2 \geq 0$, and $p_1p_2 = p_2p_1$, while it is evident from Postulates B, C that the distance of each two points of the space does not exceed πr . If the distance of two points of the space equals πr , the points are said to be diametral.

A set of points which is congruent § or isometric with a subset of $S_{n,r}$ is called r -spheric (S_n). It is clear that every three points of S_r is isometric with three points of $S_{2,r}$.

LEMMA 2.1. *Let p, q be any two non-diametral points of S_r . There is at most one point x of S_r such that $px + xq = pq$ and $px = c \cdot pq$, c , a constant, $0 < c < 1$.*

† I wish to express my thanks to Dr. Kurt Gödel for numerous suggestions given me during the preparation of this paper.

‡ This postulate may be stated arithmetically in terms of the vanishing of the determinant Δ of the four points provided any one of its third-order principal minors vanish.

The space S_r may also be defined as a *semi-metric* space such that (i) no distance exceeds πr , (ii) the determinant Δ formed for each triple of points is non-negative, and (iii) each set of four points of the space satisfies Postulate D. This definition of S_r does not involve explicitly the triangle inequality.

§ Two semi-metric sets M, M' are congruent or isometric if they may be mapped isometrically upon each other. We write $M \approx M'$.

Proof. Suppose x, x^* are two points of S_r which satisfy the conditions of the lemma, and consider the four points p, x, x^*, q . The triple p, x, q is clearly isometric with three points of $S_{1,r}$ and hence, by Postulate D there are four points $p', x', x^{*'}, q'$ of $S_{2,r}$ isometric with the points p, x, x^*, q . But, by hypothesis x and x^* are each between $\dagger p$ and q ; hence x' and $x^{*'}$ are each between p' and q' and the points $p', x', x^{*'}, q'$ lie on $S_{1,r}$, since $p'q' = pq < \pi r$. Evidently $x' = x^{*'}$; then $x = x^*$ and the lemma is proved.

A semi-metric space is called *convex* provided that for each two distinct points of the space, the space contains at least one point between them. Now Menger has shown that if p, q are any two distinct points of a convex, complete, metric space R , then R contains at least one point x that satisfies the conditions of Lemma 2. 1. Since S_r is a metric space we have, combining Lemma 2. 1 and the theorem cited above, the following theorem:

THEOREM 2. 1. *If p, q are any two distinct non-diametral points of S_r and if S_r is convex and complete, then there is exactly one point x of S_r such that $px + xq = pq$ and $px = c \cdot pq$, c , a constant, $0 < c < 1$.*

Letting c take on all values between zero and one, we call the set of points $p + q + \{x\}$, the arc pq . Then evidently two distinct non-diametral points of the convex, complete space S_r determine exactly one arc joining them, and this arc is congruent with an arc of length pq of the circle $S_{1,r}$. The arc consists of the two points together with all the points of S_r between them. If two points of S_r are diametral, every point of S_r distinct from them lies between them and hence, in general, two such points do not determine an arc.

We suppose, further, that to each point p of S_r there corresponds a point \bar{p} of S_r diametral to p ; that is, $p\bar{p} = \pi r$. A space S_r having this property is said to be *diameterized*. It is clear from Postulate C that S_r can contain at most one point \bar{p} diametral to p . Consider now two non-diametral points p, q of S_r and let \bar{p}, \bar{q} be the corresponding diametral points of p, q respectively. Then the pairs of points $\bar{p}, q; \bar{p}, \bar{q}; \bar{q}, p$ are non-diametral and hence they determine three arcs. The set of points $C_r = \text{arc } pq + \text{arc } \bar{p}q + \text{arc } \bar{p}\bar{q} + \text{arc } p\bar{q}$ is called a *great circle* of S_r . A great circle of S_r is evidently determined by two non-diametral points of S_r .

We may now prove the following theorem:

THEOREM 2. 2. *A great circle C_r of S_r is congruent to the $S_{1,r}$.*

Proof. Let p, q be two non-diametral points that determine the great circle C_r . Since $pq < \pi r$ we have $p, q, \bar{p}, \bar{q} \sim p', q', \bar{p}', \bar{q}'$, where the

\dagger A point r lies between two points p, q provided $p \neq r, r \neq q, pr + rq = pq$. The triple is said to be *linear* in the order prq .

"primed" points are points of $S_{1,r}$. We have $\text{arc } pq \approx \text{arc } p'q'$, $\text{arc } \bar{p}q \approx \text{arc } \bar{p}'q'$, $\text{arc } p\bar{q} \approx \text{arc } p'\bar{q}'$, $\text{arc } \bar{p}\bar{q} \approx \text{arc } \bar{p}'\bar{q}'$. If now, x is any point of C_r then x is a point of one of the four arcs forming C_r . Let x' of $S_{1,r}$ be the point corresponding to x by virtue of the proper one of the above congruences. This gives a mapping of C_r upon $S_{1,r}$. To show this mapping is a congruent one let $x' \sim x$, $y' \sim y$. We wish to show that $xy = x'y'$.

We observe, first, that if x and y are both points of one of the four arcs, then $xy = x'y'$ by one of the above congruences of arcs. Suppose that x, y are not both points of the same arc and, to fix the ideas, let x be an interior point† of the arc pq and y an interior point of arc $\bar{p}q$. We have $qy + y\bar{p} = q\bar{p}$, $pq + q\bar{p} = p\bar{p}$, $p\bar{p} = py + y\bar{p}$. Adding these equalities, we obtain $pq + qy = py$. Since $p, x, q \approx p', x', q'$ of $S_{1,r}$, by Postulate D the four points p, q, x, y are congruent to four points p'', q'', x'', y'' of $S_{2,r}$, and since p, q, y are linear, the four points are on $S_{1,r}$. We may make a congruent transformation of $S_{1,r}$ into itself so that $p'' \rightarrow p', q'' \rightarrow q'$, and $x'' \rightarrow x'$. This transformation sends y'' into a uniquely determined point y^* , and we have $p, q, x, y \approx p', q', x', y^*$.

Now, $q'y' + y'\bar{p}' = q'\bar{p}'$, $p'q' + q'\bar{p}' = p'\bar{p}'$, $p'\bar{p}' = p'y' + y'\bar{p}'$; adding, we obtain $p'q' + q'y' = p'y'$. From the above congruence and the preceding, $p'q' + q'y^* = p'y^*$, while $q'y' = qy = q'y^*$. Hence $p'y' = p'y^*$ and $y^* = y'$. Then $p, q, x, y \approx p', q', x', y'$ and $xy = x'y'$.

A space which is such that every quadruple of its points is congruent with some four points of the three-dimensional spherical space $S_{3,r}$ is said to have the spherical four-point property.

THEOREM 2.3. *If the space S_r is convex, complete and diameterized, the space has the spherical four-point property.*

Proof. Let p_1, p_2, p_3, p_4 be four points of S_r . If the points contain a triple congruent with three points of $S_{1,r}$, then, by Postulate D, the four points are congruent with four points of $S_{2,r}$, and hence, in this case, the theorem is proved.

We suppose, now, that no triple contained in the four points is congruent with a triple of $S_{1,r}$. By Postulates A, B, C, $p_1, p_3, p_4 \approx p'_1, p'_3, p'_4$ of $S_{2,r}$, and $p_2, p_3, p_4 \approx p''_2, p''_3, p''_4$ of $S_{2,r}$. Since $p'_3p'_4 = p_3p_4 = p''_3p''_4$, we may transform $S_{2,r}$ congruently into itself so that $p''_3 \rightarrow p'_3, p''_4 \rightarrow p'_4$. The point p_2 is transformed into one of two points p_2^I, p_2^{II} which are images of each other in the great circle of $S_{2,r}$ determined by p'_3, p'_4 . Denote this circle by $S_{1,r}(p'_3, p'_4)$. We have $p_2, p_3, p_4 \approx p_2^I, p'_3, p'_4 \approx p_2^{II}, p'_3, p'_4$. Now p'_1 is not on $S_{1,r}(p'_3, p'_4)$, so we may assume the labelling so that p'_1, p_2^I are both

† If x, y are not interior points of their respective arcs, the procedure is simplified.

on one of the two hemi-spheres determined by $S_{1,r}(p'_3, p'_4)$. Then $p'_1 p_2^I < p'_1 p_2^{II}$.

Let x' denote the intersection of the arc $p'_1 p_2^{II}$ with $S_{1,r}(p'_3, p'_4)$ †. The great circle $C_r(p_3, p_4)$ of S_r determined by p_3, p_4 is congruent to $S_{1,r}(p'_3, p'_4)$. In this congruence, let x of S_r correspond to x' . By Postulate D , $p_1, p_3, p_4, x \approx \bar{p}_1, \bar{p}_3, \bar{p}_4, \bar{x}$ of $S_{2,r}$. Since $\bar{p}_1, \bar{p}_3, \bar{p}_4 \approx p_1, p_3, p_4 \approx p'_1, p'_3, p'_4$ of $S_{2,r}$, not $S_{1,r}$, a congruent transformation of $S_{2,r}$ into itself gives $p_1, p_3, p_4, x \approx p'_1, p'_3, p'_4, x^*$, with x^* uniquely determined. But $p'_3, p'_4, x' \approx p_3, p_4, x \approx p'_3, p'_4, x^*$, and p'_3, p'_4 are not diametral. Hence $x^* = x'$ and we have $p_1, p_3, p_4, x \approx p'_1, p'_3, p'_4, x'$. In a similar manner, we obtain $p_2, p_3, p_4, x \approx p_2^{II}, p_3, p_4, x'$. Now $p_1 p_2 \leq p_1 x + x p_2 = p'_1 x' + x' p_2^{II} = p'_1 p_2^{II}$.

Let the arc $p'_1 p_2^I$, extended, meet $S_{1,r}(p'_3, p'_4)$ in a point y' . Let y of S_r correspond to y' in the congruence $C_r(p_3, p_4) \approx S_{1,r}(p'_3, p'_4)$. As above, we obtain $p_1, p_3, p_4, y \approx p'_1, p'_3, p'_4, y'$, and $p_2, p_3, p_4, y \approx p_2^I, p'_3, p'_4, y'$. Now $p_1 p_2 \geq |p_1 y - p_2 y| = |p'_1 y' - p_2^I y'| = p'_1 p_2^I$. Hence $p'_1 p_2^I \leq p_1 p_2 \leq p'_1 p_2^{II}$.

Rotate the $S_{3,r}$ containing the $S_{2,r}$ half a revolution about $S_{1,r}(p'_3, p'_4)$ as axis. The point p_2^I is carried by this rotation into the point p_2^{II} . Let p'_x denote any point of its path. Then the function $p'_1 p_x$ is a continuous function which takes on the values $p'_1 p_2^I$ and $p'_1 p_2^{II}$, and consequently all values between these two. But $p_1 p_2$ has been shown to lie between $p'_1 p_2^I$ and $p'_1 p_2^{II}$. Hence, the $S_{3,r}$ contains a point, say p'_2 , such that $p'_1 p'_2 = p_1 p_2$, and p'_2 has the same distances from p'_3, p'_4 as p_2^I or p_2^{II} . Hence $p_1, p_2, p_3, p_4 \approx p'_1, p'_2, p'_3, p'_4$ of $S_{3,r}$, and the theorem is proved.‡

THEOREM 2.4. *Let C_r^1, C_r^2 be two distinct great circles of S_r having a point p in common. Then the set of points $C_r^1 + C_r^2$ can be isometrically imbedded in $S_{2,r}$.*

Proof. Let p_1, p_2 be points of C_r^1, C_r^2 respectively, distinct from p and not diametral to p . By Postulates B, C we have $p, p_1, p_2 \approx p', p'_1, p'_2$ of $S_{2,r}$. Let $S_{1,r}^1, S_{1,r}^2$ denote the great circles of $S_{2,r}$ determined by the pairs $p', p'_1; p', p'_2$, respectively. By the preceding theorem, $C_r^1 \approx S_{1,r}^1, C_r^2 \approx S_{1,r}^2$. These two congruences give a mapping of the set $C_r^1 + C_r^2$ upon the set $S_{1,r}^1 + S_{1,r}^2$. We must show this mapping is a congruent one.

Let x, y be any two points of $C_r^1 + C_r^2$ and suppose them mapped into x', y' respectively, by means of the above mapping. We show that $xy = x'y'$.

† If p'_1, p_2^{II} are diametral, any point on $S_{1,r}(p'_3, p'_4)$ may be taken as the point x' .

‡ In the proof given above of Theorem 2.3 the assumption that the space S_r is diameterized is needed to insure the existence in S_r of the great circle $C_r(p_3, p_4)$, and hence of the points x, y corresponding to x', y' , respectively, of $S_{1,r}(p'_3, p'_4)$. A slightly different procedure, however, found by the writer while this paper was in press, establishes the theorem *without this assumption*. This applies also to Theorem 3.3 of the following section.

If x, y are both points of one of the two circles in S_r , then $xy = x'y'$ by virtue of one of the above congruences. Suppose that x is a point of C_r^1 and y a point of C_r^2 . By Postulate D we have $p, x, p_1, p_2 \approx p'', x'', p'_1, p''_2$ of $S_{2,r}$. Now $p', p'_1, p'_2 \approx p, p_1, p_2 \approx p'', p''_1, p''_2$, and hence we may transform $S_{2,r}$ congruently into itself so that $p'' \rightarrow p', p''_1 \rightarrow p'_1, p''_2 \rightarrow p'_2$. Since p, p_1, p_2 are evidently not on a great circle of S_r , the points p', p'_1, p'_2 are not on a great circle of $S_{2,r}$ and consequently the above congruent transformation sends x into a uniquely determined point x^* . We have $p, x, p_1, p_2 \approx p', x^*, p'_1, p'_2$ in $S_{2,r}$. Now, $p', x', p'_1 \approx p, x, p_1 \approx p', x^*, p'_1$ on $S_{1,r}^1$. Hence $x^* = x'$, and $p, x, p_1, p_2 \approx p', x', p'_1, p'_2$.

Consider, now, the quadruple p, p_2, x, y . We obtain by the usual procedure, $p, p_2, x, y \approx p', p'_2, x', y^*$ on $S_{2,r}$. As above, $p', p'_2, y' \approx p, p_2, y \approx p', p'_2, y^*$ on $S_{1,r}^2$. Hence $y^* = y'$ and $p, p_2, x, y \approx p', p'_2, x', y'$. Then $xy = x'y'$.

Using the result of Theorem 2.3 it is readily shown that if three great circles of S_r have a point in common, their sum may be imbedded congruently in $S_{3,r}$.

THEOREM 2.5. *Let p_1, p_2, p_3 be three points of the convex, complete space S_r which are not on a great circle of S_r . If we denote the three arcs they determine by A_{12}, A_{23}, A_{13} , then the set $A_{12} + A_{23} + A_{13}$ can be isometrically imbedded in $S_{2,r}$.*

The proof of this theorem is immediate. We remark that if the great circles $C_r^{12}, C_r^{23}, C_r^{13}$ exist in S_r , their sum may be imbedded in $S_{2,r}$. Since in Theorem 2.5 we are not assuming that S_r is diameterized, these great circles may not exist.

3. We define n -dimensional sub-spaces $\Omega_{n,r}$ of S_r , assumed convex, complete, and diameterized, by recursion. Call a single point a zero-dimensional sub-space $\Omega_{0,r}$ of S_r and consider $n+1$ points p_0, p_1, \dots, p_n of S_r which are congruent to $n+1$ points of $S_{n,r}$ not of $S_{n-1,r}$. (If $n=1$, we assume that the two points are not diametral). If the points p_1, p_2, \dots, p_n determine an $(n-1)$ -dimensional sub-space $\Omega_{n-1,r}$ of S_r , the union of the great circles $\{p_0, x\}$, as x ranges over $\Omega_{n-1,r}$ is defined as an n -dimensional sub-space $\Omega_{n,r}$ of S_r with base points p_0, p_1, \dots, p_n . Thus, for $n=1$, $\Omega_{1,r}$ is the great circle of S_r determined by p_0, p_1 which has been shown to exist, is unique, and is congruent with the great circle $S_{1,r}$ of $S_{n,r}$ determined by p'_0, p'_1 where $p_0, p_1 \approx p'_0, p'_1$. It is obvious that no point of $\Omega_{n-1,r}$ is diametral with p_0 .

This definition assigns a special rôle to the point p_0 . It will be shown, however, that the base points uniquely determine the n -dimensional sub-space which is, therefore, independent of the particular vertex "isolated."

THEOREM 3.1. *Let p_0, p_1, p_2 be three points of the convex, complete diameterized space S_r , congruent with p'_0, p'_1, p'_2 , three points of $S_{2,r}$, not of $S_{1,r}$. Then these three points determine exactly one two-dimensional sub-space $\Omega_{2,r}$ of S_r , with base points p_0, p_1, p_2 , and this two dimensional sub-space is congruent with the two-dimensional spherical sub-space $S_{2,r}$ of $S_{n,r}$ determined by the three points p'_0, p'_1, p'_2 . Moreover, the congruence of $\Omega_{2,r}$ with $S_{2,r}$ containing the congruence $p_0, p_1, p_2 \approx p'_0, p'_1, p'_2$ is unique.*

Proof. Let x be a point of $\Omega_{2,r}$. Then the great circle $C_r(p_1, p_2)$ of S_r contains a point σ such that x is a point of the great circle $C_r(p_0, \sigma)$ of S_r . By Theorem 2.2, $C_r(p_0, \sigma) \approx S_{1,r}(p'_0, \sigma')$.† Let x' of $S_{2,r}$ correspond to x by means of this congruence. This gives a mapping of $\Omega_{2,r}$ upon a sub-set of $S_{2,r}$.

If y is a point of $\Omega_{2,r}$ distinct from x , let y' of $S_{2,r}$ correspond to y by the above mapping. If y is contained in $C_r(p_0, \sigma)$ then $xy = x'y'$ by the congruence of great circles. If y is not a point of $C_r(p_0, \sigma)$ then there exists a point τ of $C_r(p_1, p_2)$ distinct from σ and $\bar{\sigma}$ such that y belongs to $C_r(p_0, \tau)$. Then p_0, σ, τ are three points of S_r which are not on a great circle of S_r and hence, by the remark following Theorem 2.5 the sum of the three great circles of S_r they determine may be isometrically embedded in $S_{2,r}$. This yields $xy = x'y'$ and the mapping is a congruent one.

Denote by $\Omega'_{2,r}$ the subset of $S_{2,r}$ into which $\Omega_{2,r}$ has been congruently transformed and let x' be any point of $S_{2,r}$. Now $p_0, p_1, p_2 \approx p'_0, p'_1, p'_2$ of $S_{2,r}$ not $S_{1,r}$, and x' is a point of a great circle of $S_{2,r}$ that contains p'_0 and a point s' of the great circle $S_{1,r}(p'_1, p'_2)$. Let s in S_r correspond to s' by virtue of the congruence $C_r(p_1, p_2) \approx S_{1,r}(p'_1, p'_2)$. Let x of S_r be selected on the great circle $C_r(p_0, s)$ so that $p_0, s, x \approx p'_0, s', x'$. Then x is a point of $\Omega_{2,r}$. Now since $\Omega_{2,r} \approx \Omega'_{2,r}$ there is a point x^* of $\Omega'_{2,r}$ that corresponds to x of $\Omega_{2,r}$ in this congruence. Evidently $x^* = x'$. Then x' is a point of $\Omega'_{2,r}$ and $\Omega'_{2,r} = S_{2,r}$.

To show that there is but a single two-dimensional sub-space with the points p_0, p_1, p_2 as base points, denote by $\tilde{\Omega}_{2,r}$ any other such space formed from these points. We have (1) $\Omega_{2,r} \approx S_{2,r}$ and (2) $\tilde{\Omega}_{2,r} \approx S_{2,r}$. Suppose x' is any point of $S_{2,r}$, distinct from p'_0, p'_1, p'_2 and let x of $\Omega_{2,r}$ \tilde{x} of $\tilde{\Omega}_{2,r}$ be corresponding points by means of congruences (1), (2) respectively. Let x' be on the arc p'_0s' , where s' is a point of $S_{1,r}(p'_1, p'_2)$ and let s of $\Omega_{2,r}$, \bar{s} of $\tilde{\Omega}_{2,r}$ correspond to s' in the appropriate congruences. Then

$$p_1, p_2, s \approx p'_1, p'_2, s' \approx p_1, p_2, \bar{s} \quad \cdot \cdot$$

† We have $C_r(p_1, p_2) \approx S_{1,r}(p'_1, p'_2)$ and $\sigma' \sim \sigma$ by this congruence.

where s, \bar{s} are on $C_r(p_1, p_2)$. Then $s = \bar{s}$, and it follows immediately that $x = \bar{x}$, for

$$p_0, s, x \approx p'_0, s', x' \approx p_0, \bar{s}, \bar{x} \approx p_0, s, \bar{x}.$$

Finally, we observe that the congruence $\Omega_{2,r} \approx S_{2,r}$ containing the congruence $p_0, p_1, p_2 \approx p'_0, p'_1, p'_2$ is unique, for the contrary implies a congruent transformation of $S_{2,r}$ into itself leaving the points p_0, p_1, p_2 fixed and which is not an identical transformation. This, of course, is impossible.

The analogous theorem concerning the n -dimensional sub-space is proved by complete induction. The procedure employed in the proof of Theorem 3.1 is easily generalized and the following theorem is obtained.

THEOREM 3.2. *Let p_i ($i=0, 1, \dots, n$) be $n+1$ points of the complete, complex, diameterized space S_r which are congruent with $n+1$ points p'_i ($i=0, 1, \dots, n$) of $S_{n,r}$, not of $S_{n-1,r}$. There is exactly one n -dimensional sub-space $\Omega_{n,r}$ of S_r that has the points p_i ($i=0, 1, \dots, n$) as base points and the congruence $p_0, p_1, \dots, p_n \approx p'_0, p'_1, \dots, p'_n$ induces a unique congruence of $\Omega_{n,r}$ with the $S_{n,r}$.*

The existence of the sub-space $\Omega_{n,r}$ congruent with $S_{n,r}$ depends on the space S_r containing $n+1$ points that may be imbedded isometrically in the $S_{n,r}$. By Theorem 2.3, every four points of S_r , assumed convex, complete and diameterized, is congruent to four points of $S_{3,r}$; that is S_r has the spherical four-point property. We show, by complete induction that this implies that such a space has the spherical n -point property for every integer n .

THEOREM 3.3. *The convex, complete, diameterized space S_r has the spherical n -point property for every integer n .*

Proof. We suppose that S_r has the spherical k -point property for every $k \leq n$, where n is a fixed integer greater than three. Let p_0, p_1, \dots, p_n be any $n+1$ points of S_r . By hypothesis, any k of these points, $k \leq n$, are congruent with k points of $S_{k-1,r}$. Now either the given $n+1$ points are congruent with $n+1$ points of $S_{n-1,r}$ (in which case the theorem is proved) or n points of the set may be selected to form the base points of an $(n-1)$ -dimensional subspace $\Omega_{n-1,r}$ of S_r which does not contain the remaining point.

In this case we assume the labelling so that

$$p_0, p_2, \dots, p_n \approx p'_0, p'_2, \dots, p'_n \text{ of } S_{n-1,r},$$

not of $S_{n-2,r}$, and

$$p_1, p_2, \dots, p_n \approx \bar{p}_1, \bar{p}_2, \dots, \bar{p}_n \text{ of } S_{n-1,r},$$

since the spherical k -point property is assumed for $k \leq n$. We may transform $S_{n-1,r}$ congruently into itself so that $\bar{p}_2 \rightarrow p'_2, \dots, \bar{p}_n \rightarrow p'_n$. This transforms

\bar{p}_1 into a point p_1^* , which has its distances from the $n-1$ points p'_2, \dots, p'_n determined. Since the points p'_2, p'_3, \dots, p'_n are not on $S_{n-2,r}$ there are *exactly two distinct positions* for the point p_1^* on the $S_{n-1,r}$ containing p_1^*, p'_2, \dots, p'_n . We denote the two positions of p_1^* by p_1^I, p_1^{II} ; these two points are evidently images of each other in the $S_{n-2,r}$ determined by p'_2, p'_3, \dots, p'_n . We have

$$p_1, p_2, \dots, p_n \approx p_1^I, p'_2, \dots, p'_n \approx p_1^{II}, p'_2, \dots, p'_n.$$

Since p'_0 does not lie on the $S_{n-2,r}$ containing p'_2, \dots, p'_n , $p'_0 p_1^I \neq p'_0 p_1^{II}$. Assume the labelling so that p'_0, p_1^I are *both* on one of the two "hemi-spheres" of $S_{n-1,r}$ determined by $S_{n-2,r}(p'_2, \dots, p'_n)$. Then $p_0 p_1^I < p_0 p_1^{II}$.

The arc $\dagger p'_0 p_1^{II}$ intersects $S_{n-2,r}(p'_2, \dots, p'_n)$ in a point x' . The $(n-2)$ -dimensional space $\Omega_{n-2,r}$ with base points p_2, p_3, \dots, p_n of S_r is congruent to $S_{n-2,r}(p'_2, p'_3, \dots, p'_n)$. In this congruence let x of S_r correspond to x' . This congruence is contained in both of the congruences

- (1) $\Omega_{n-1,r}(p_0, p_2, \dots, p_n) \approx S_{n-1,r}(p'_0, p'_2, \dots, p'_n)$
- (2) $\Omega_{n-1,r}(p_1, p_2, \dots, p_n) \approx S_{n-1,r}(p_1^I, p'_2, \dots, p'_n) \approx S_{n-1,r}(p_1^{II}, p'_2, \dots, p'_n)$.

By congruence (1), $p_0 x = p'_0 x'$ and by congruence (2), $x p_1 = x' p_1^{II}$. Now, by the triangle inequality, $p_0 p_1 \leq p_0 x + x p_1 = p'_0 x' + x' p_1^{II} = p'_0 p_1^{II}$. Again, the arc $p'_0 p_1^I$ extended, intersects $S_{n-2,r}(p'_2, \dots, p'_n)$ in a point y' . We have, as above, $p_0 y = p'_0 y'$, $p_1 y = p_1^I y'$ where y of S_r corresponds to y' by the congruence $\Omega_{n-2,r}(p_2, \dots, p_n) \approx S_{n-2,r}(p'_2, \dots, p'_n)$. Then

$$p_0 p_1 \geq |p_0 y - p_1 y| = |p'_0 y' - p_1^I y'| = p'_0 p_1^I.$$

Hence

$$p'_0 p_1^I \leq p_0 p_1 \leq p'_0 p_1^{II}.$$

Suppose the $S_{n,r}$ containing the $S_{n-1,r}(p'_0, p_1^I, p_1^{II}, p'_2, \dots, p'_n)$ is rotated 180 degrees about the $S_{n-2,r}(p'_2, \dots, p'_n)$ as axis. This rotation keeps fixed the points p'_2, \dots, p'_n , while the point p_1^I is carried over to the point p_1^{II} . Let p'_x denote any point of its path. Keeping the label of point p'_0 fixed, the function $p'_0 p'_x$ is a continuous function which assumes the values $p'_0 p_1^I$ and $p'_0 p_1^{II}$ and hence for any value between these two, such as $p_0 p_1$, there is a point, say p'_1 , of $S_{n,r}$, such that $p'_0 p'_1 = p_0 p_1$. Hence

$$p_0, p_1, p_2, \dots, p_n \approx p'_0, p'_1, p'_2, \dots, p'_n \text{ of } S_{n,r}$$

and the theorem is proved.†

4. Let 0 be a point of Hilbert space. The set $\{x\}$ of points of the

† If $p'_0 p_1^{II}$ are diametral, any point of $S_{n-2}(p'_2, \dots, p'_n)$ may be taken as the point x' .

‡ If $p_0 p_1$ equals $p'_0 p_1^I$ or $p'_0 p_1^{II}$ the $n+1$ points are evidently congruent with $n+1$ points of $S_{n-1,r}$.

space whose distance from 0 is equal to a positive constant r is called the sphere with center 0 and radius r in Hilbert space, and is denoted by H_r .

Let p_i, p_j be any two points of S_r and denote by α_{ij} the angle $p_i p_j / r$ radians. If p_1, p_2, \dots, p_k are any k points of S_r , we denote the axisymmetric determinant

$$\begin{vmatrix} 1 & \cos \alpha_{12} & \cdots & \cos \alpha_{1k} \\ \cos \alpha_{21} & 1 & \cdots & \cos \alpha_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \cos \alpha_{k1} & \cos \alpha_{k2} & \cdots & 1 \end{vmatrix}$$

by $\Delta_k(p_1, p_2, \dots, p_k)$.

If the space S_r is such that there exists an integer k such that the determinant Δ_k formed for every set of k points of S_r vanishes, we say the space S_r is *indexed*. If S_r is indexed there is evidently a smallest integer for which S_r has this property. Call this integer the index of S_r .

We now prove the fundamental theorem characterizing metrically the n -dimensional spherical space $S_{n,r}$.

THEOREM 4.1. *The convex, complete, diameterized and indexed space S_r with index $n+2$ is congruent with $S_{n,r}$.*

Proof. Since S_r has index $n+2$ there exists at least one set of $n+1$ points of S_r for which Δ_{n+1} does not vanish, while $\Delta_{n+2} = 0$ for every set of $n+2$ points of S_r . By Theorem 3.3, every set of $n+1$ points of S_r is congruent with $n+1$ points of $S_{n,r}$. Hence, since Δ_{n+2} vanishes, every set of $n+2$ points of S_r is congruent with $n+2$ points of $S_{n,r}$.†

Consider, now, any set of $n+3$ points of S_r . Since each set of $n+2$ points contained in these $n+3$ points may be imbedded isometrically in $S_{n,r}$, the $n+3$ points are either congruent with $n+3$ points of $S_{n,r}$ or they form a pseudo r -spheric $(n+3)$ -tuple. But each set of $n+3$ points of S_r is, by Theorem 3.3, congruent with $n+3$ points of $S_{n+2,r}$; that is, $\Delta_{n+3} \geq 0$; while the determinant Δ_{n+3} formed for a pseudo r -spheric $(n+3)$ -tuple is negative.‡ Hence no set of $n+3$ points of S_r can be pseudo r -spheric and therefore each $(n+3)$ -tuple is congruent with an $(n+3)$ -tuple of $S_{n,r}$. Since S_n has the congruence order $n+3$, this implies that S_r is congruent to a subset of $S_{n,r}$.

Now S_r contains a set of $n+1$ points for which Δ_{n+1} does not vanish. These points are, therefore, not congruent with $n+1$ points of $S_{n-1,r}$ and

† See Theorem II_n of Blumenthal and Garrett, "Characterization of spherical and pseudo-spherical sets of points," *American Journal of Mathematics*, Vol. 55 (1933), p. 632.

‡ *Loc. cit.*, Theorems I_n , IV_n , p. 632.

hence they determine a sub-space $\Omega_{n,r}$ of S_r which is congruent to $S_{n,r}$. Hence $S_r \approx S_{n,r}$ and the theorem is proved.

The metric characterization of the sphere H_r in Hilbert space is given by the following theorem.

THEOREM 4.2. *If the convex, complete, diameterized, separable space S_r is not indexed it is congruent to H_r , the sphere in Hilbert space.*

Proof. By Theorem 3.3, every k points of S_r is congruent with k points of $S_{k-1,r}$; that is, $\Delta_k \geq 0$ for every integer k . Since S_r is separable, we may conclude that S_r is congruent to a subset of H_r .†

Since S_r is not indexed, for every integer k , S_r contains at least one set of k points for which Δ_k does not vanish and hence is positive. Then S_r is not isometric with $S_{n,r}$ for any value of n .

It follows readily from the completeness of S_r that $S_r \approx H_r$.

If each of the $n+3$ sets of $n+2$ points contained in $n+3$ points of a semi-metric space is congruent with $n+2$ points of $S_{n,r}$, while the $n+3$ points are not congruent with $n+3$ points of $S_{n,r}$, the set is said to form a pseudo r -spheric $(n+3)$ -tuple. From the foregoing theorems it follows that a convex, complete, diameterized space S_r cannot contain pseudo r -spheric sets. Thus, for example, the set of five points p_1, p_2, \dots, p_5 , with the distance of each pair defined to be $r \cdot \cos^{-1}(-1/3)$, has each of its quadruples isometric with four points of $S_{2,r}$, while the five points are clearly not isometric with five points of $S_{2,r}$. As a consequence of our theorems these five points cannot be imbedded in a space S_r which is convex, complete, and diameterized.

It may be observed that the theorems obtained by W. A. Wilson in his investigation of metric spaces in which each four points are congruent with four points of a euclidean space, are limiting cases of the theorems presented in this paper, while Postulate D is weaker than the four-point condition assumed by Wilson.‡

THE INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

† Menger has shown (*Ergebnisse eines Mathematischen Kolloquiums*, Wien, Heft 1 (1931), p. 26) that a necessary and sufficient condition for a separable semi-metric space R to be congruent with a subset of Hilbert space is that for every integer k and for every k points p_1, p_2, \dots, p_k of R , $\text{sign } D(p_1, p_2, \dots, p_k) = \text{sign } (-1)^k$ or 0, where $D(p_1, p_2, \dots, p_k)$ is the well-known volume-determinant. We obtain the result above by replacing $p_i p_j$ by $2r \cdot \sin p_i p_j / 2r$ in Menger's theorem.

It may be pointed out that there is a typographical error in the statement of this theorem occurring in Menger, "Die Metrik des Hilbert'schen Raumes," *Wiener Akademischer Anzeiger*, 1928, No. 12, p. 160.

‡ W. A. Wilson, "A relation between metric and euclidean spaces," *American Journal of Mathematics*, Vol. 54 (1932), pp. 505-517.

ON CERTAIN TYPES OF CONTINUOUS TRANSFORMATIONS OF METRIC SPACES.*

By W. A. WILSON.

1. The literature on topology produced during the last two decades has by its great volume tended to divert attention from the study of geometrical properties which are not invariant under continuous transformations. It may well be that with more restricted transformations the methods used in topological problems can be applied with equal success to the study of non-topological properties of geometrical configurations. It is the purpose of this article to develop some of the characteristics of a certain class of continuous transformations and their application to geometry,—with the hope that they indicate the possibility of further progress in the direction mentioned above.

2. *Definition of spread at a point.* Let $X = \{x\}$ be a metric space and, as usual, let xx' denote the distance between the points x and x' . If $Y = \{y\}$ is another metric space which is the continuous image of X by a transformation $y = f(x)$, one-valued and defined over X , there are three possibilities at any limiting point a of X . First, there may be a constant $m_f(a)$ such that, if y and y' correspond to distinct points x and x' , respectively, $\lim(yy'/xx') = m_f(a)$ as x and x' approach a . (It is understood that either x or x' may coincide with a .) In this case we say that the transformation has a *finite spread at a* and denote the spread by $m_f(a)$. Second, it may happen that $\lim(yy'/xx') = \infty$; in this case we say that the *spread is infinite at a* and write $m_f(a) = \infty$. Finally, yy'/xx' may approach no limit as x and x' approach a , and we then say that the *spread is indeterminate at a* .

That the last case can arise effectively is seen in a homeomorphism between a straight line and a broken line. Examples of the first two are easily invented. We shall be chiefly concerned with transformations for which the spread is defined at each point and is constant.

3. *Elementary properties of the spread.* From the definition it is apparent that the spread is somewhat analogous to the derivative. In fact, it is easy to show that, if x and y are real variables and $y = f(x)$ is continuous and differentiable, continuity of $f'(x)$ at $x = a$ implies that $m_f(a)$ is finite, and conversely; moreover, $m_f(a) = |f'(a)|$. As in the case of derivatives, the following properties are deducible from the definition:

* Presented to the Society, September, 1934.

I. Let $y=f(x)$ and its inverse $x=g(y)$ define a homeomorphism between the metric spaces X and Y . If $b=f(a)$ and $m_f(a)$ exists, finite or infinite, so does $m_g(b)$. If $m_f(a) \neq 0$ and is finite, $m_g(b) = 1/m_f(a)$; if $m_f(a) = 0$, $m_g(b) = \infty$; and if $m_f(a) = \infty$, $m_g(b) = 0$.

II. Let $X=\{x\}$ be a compact metric space and the continuous transformation $y=f(x)$ have a finite spread at each limiting point of X . Then yy'/xx' is bounded in X .

III. Let $X=\{x\}$ be a compact metric space and the transformation $y=f(x)$ have a finite spread $m_f(x)$ which is continuous in X . Then yy'/xx' converges to $m_f(x)$ uniformly in X .

The proof of I is like that of the corresponding theorem on derivatives. If we assume that II fails, it follows from the compactness of X and the continuity of $f(x)$ that there is some point a and points x and x' converging to a for which $yy'/xx' \rightarrow \infty$, in contradiction to the hypothesis that $m_f(a)$ is finite. If III fails, there is a sequence of points $\{a_i\}$ converging to a point a and points x_i and x'_i such that $a_ix_i \rightarrow 0$ and $a_ix'_i \rightarrow 0$, but $|m_f(a_i) - y_iy'_i/x_ix'_i|$ exceeds a preassigned positive ϵ . This contradicts the hypothesis that $m_f(a_i) \rightarrow m_f(a)$.

4. *Applications to rectifiability.* Let us suppose that $y=f(x)$ defines a homeomorphism between the metric spaces X and Y . If the spread is finite at each point, it is evident by Property II of the previous section and the usual definition of length of arc, that rectifiable arcs in X go over into rectifiable arcs in Y . If the spread is continuous and for some arc in X the limit of the ratio of the length of a sub-arc ax to the distance ax is any constant m , as x approaches a , this property is also preserved in the image by virtue of Property III. When the spread is a constant $k \neq 0$, the length of the image of an arc of length s is easily seen to be ks . (In this case, if intrinsic distances* are used, the transformation is one of similitude.) If the spread is everywhere infinite, it follows from Properties I and III that the image of no rectifiable arc is rectifiable. If the spread is everywhere zero, X can contain no rectifiable arc, since the image of such an arc would be a point.

On the other hand we can have a homeomorphic transformation with a constant zero spread of a metric space which contains no rectifiable arc, as shown by the following example. Let $Y=\{y\}$ be the linear interval $0 \leq y \leq 1$. Let $X=\{x\}$ be a metric space obtained from Y by the con-

* Compare the writer's paper, "On rectifiability in metric spaces," *Bulletin of the American Mathematical Society*, vol. 38, pp. 419-426.

tinuous transformation $x = g(y)$ such that, for any two points y, y' and the corresponding points x, x' , we have $xx' = \sqrt{yy'}$. It is easily seen that this transformation is a homeomorphism and that X is a simple arc. If $y = f(x)$ is the inverse of $x = g(y)$, we see that $yy'/xx' = xx'$; hence $m_f(x) = 0$ at every point of X and $m_g(y) = \infty$ at every point of Y .

The arc X has some remarkable properties. Any $n + 1$ points a_0, a_1, \dots, a_n may be arranged so that, if $b_i = f(a_i)$, $b_i < b_{i+1}$. Setting $r_i = b_{i-1}b_i$, we have $a_0a_1 = \sqrt{r_1}$, $a_0a_2 = \sqrt{r_1 + r_2}$, $a_0a_3 = \sqrt{r_1 + r_2 + r_3}$, \dots , $a_1a_2 = \sqrt{r_2}$, etc. Substituting these values in the $(n + 1)$ -point determinant D_n used by Menger* as a criterion of imbeddability in Euclidean space, we find by easy algebra that $D_n = (-1)^{n+1} \cdot 2^n \cdot r_1 \cdot r_2 \cdot \dots \cdot r_n$. Hence the arc can be imbedded† in Hilbert space, no $n + 1$ of its points can be imbedded in an $(n - 1)$ -dimensional Euclidean space, and in particular no three of its points lie in a straight line. The curve is extremely crinkly; not only is no part of it rectifiable, but, if a, b , and c are any three points in order, abc is a right angle.

5. *Unrestricted regular transformations.* The foregoing example suggests the following definition. Let $v = \phi(u)$ be a one-valued real function defined for $u \geq 0$ and subject to these conditions: (a) $\phi(u) > 0$ if $u > 0$; (b) $\phi(0) = 0$; (c) $\phi(u)$ is continuous; (d) if $u_1 + u_2 \geq u_3$, then $\phi(u_1) + \phi(u_2) \geq \phi(u_3)$. If $y = f(x)$ is a continuous transformation of a metric space X into a metric space Y such that, if u is the distance between two points in X , the distance between the image points in Y is $v = \phi(u)$, where $\phi(u)$ satisfies Conditions (a) to (d) above, we say that the transformation is an *unrestricted regular transformation* and call $\phi(u)$ the *scale function* of the transformation.

Such a transformation is regular in the sense that the spread is constant (as will be seen soon) and congruent figures go over into congruent figures. It is unrestricted in the sense that it can be applied to any metric space. A transformation whose scale function satisfies the given conditions except for (d) would preserve congruences and might be applicable to a certain space X , but not applicable to a metric space containing X . Simple examples of functions satisfying the given requirements are $v = \sqrt{u}$, $v = \log(1 + u)$, and $v = \sin u$ for $u \leq \pi/2$ and $v = 1$ for $u > \pi/2$. The function $v = u^2$ does not satisfy Condition (d) and so a transformation involving this relation

* "Untersuchungen über allgemeine Metrik," *Mathematische Annalen*, vol. 100, pp. 119, 120.

† I. e., is congruent with a sub-set of Hilbert space.

between distances would be impossible unless for every triple of points in the given space $|(ab)^2 - (ac)^2| \leq (bc)^2 \leq (ab)^2 + (ac)^2$.

Any scale function $v = \phi(u)$ of an unrestricted regular transformation has a variety of properties,* of which the following are important:

- I. It is monotone increasing.
- II. $\phi'(0)$ exists, finite or infinite, and $\phi'(0) > 0$.
- III. For any value of u , $\limsup_{\Delta u \rightarrow 0} (\Delta \phi / \Delta u) \leq \phi'(0)$.
- IV. If the inverse, $u = \psi(v)$, is also a scale function of an unrestricted regular transformation, v/u is a constant.

Property I is easily deduced from Conditions (a) to (d). To prove Property II let M be the upper bound of $\phi(u)/u$; then $\limsup_{u \rightarrow 0} [\phi(u)/u] \leq M$. If M is finite, there is, for every $\epsilon > 0$, some value \bar{u} of u for which $\phi(\bar{u})/\bar{u} > M - \epsilon/2$. Now $\phi(u)/u$ is continuous; hence for some $\delta > 0$ and any u such that $|u - \bar{u}| < \delta$, $\phi(u)/u > M - \epsilon$. If $u' < \delta$, there is an integer n for which $|nu' - \bar{u}| < \delta$. Moreover, by repeated application of Condition (d),

$$\phi(u')/u' \geq \phi(nu')/nu' > M - \epsilon.$$

Hence $\liminf_{u \rightarrow 0} [\phi(u)/u] \geq M$. A similar argument applies if $M = \infty$. Therefore $\phi'(0)$ exists and equals M . Clearly $M \neq 0$, since $\phi(u) > 0$ if $u > 0$. Property III follows from Condition (d) in the above definition, and Property IV is a consequence of III and the fact that $\phi'(0) \cdot \psi'(0) = 1$.

THEOREM. Let $X = \{x\}$ be a dense metric space and $y = f(x)$ be an unrestricted regular transformation whose scale function is $v = \phi(u)$. Then $y = f(x)$ defines a homeomorphism and its spread is $\phi'(0)$ at each point of X .

Proof. The inverse transformation $x = g(y)$ is one-valued, since $\phi(u) > 0$ for $u > 0$. If $g(y)$ is not everywhere continuous, there is a point y_0 and a sequence $\{y_i\}$ converging to y_0 , such that, if $x_0 = g(y_0)$ and $x_i = g(y_i)$, $x_i x_0$ exceeds some positive number k for all values of i . If we let $u_i = x_i x_0$, we have $v_i = \phi(u_i) \geq \phi(k) > 0$. This is impossible, as $v_i = y_i y_0$ and $y_i \rightarrow y_0$. Hence $g(y)$ is both one-valued and continuous, and the transformation is a homeomorphism.

If a is any point of X , there are points x and x' in any vicinity of a . If we set $u = xx'$, $yy'/xx' = v/u$ and, as x and x' approach a , $u \rightarrow 0$. Hence $m_f(a) = \lim_{u \rightarrow 0} (v/u) = \phi'(0)$.

* Compare also F. Toranzos, "Über die Funktional-Ungleichung $f(x) + f(y) = f(x + y)$," *Rev. mat. hisp-amer.*, vol. 2, pp. 109-113.

6. The theorem just proved shows that the discussion of § 4 is valid for the class of unrestricted regular transformations. Furthermore, if the inverse of such a transformation is also regular and unrestricted, Property IV of scale functions shows that the transformation is one of similitude.

This is not, however, universally the case. The functions $v = \sqrt{u}$ and $v = \log(1 + u)$, mentioned earlier, are instances in point. Such transformations, when applied to ordinary spaces, produce interesting new spaces. For example, if $\phi'(0) = k \neq \infty$ and the given space is complete and convex, the image is complete and each pair of points is joined by an arc of length k times the distance between the corresponding points in the given space, but the image is not convex unless the transformation is one of similitude. If we take as the space X a unit square in the plane and apply a transformation with the scale function $v = \sqrt{u}$ (here $\phi'(0) = \infty$), the image Y is a simple two-cell, the images of segments in X of equal length being congruent simple arcs of the type described in § 4. It would be interesting to know whether this and similar configurations can be imbedded in Hilbert space.

7. *Angles and conformality.* Let A , B , and C be points in a metric space, $a = BC$, $b = AC$, and $c = AB$. Let us call the symbol BAC an *angle* with vertex A and define its value by the formula $\cos BAC = (b^2 + c^2 - a^2)/2bc$, taking the value between 0 and π . This definition is possible by virtue of the metric triangle inequality. If ρ and σ are two rays* with the common initial point A , and B and C are other points on ρ and σ , respectively, we say that ρ and σ make an angle (ρ, σ) if $\lim BAC$ exists when B and C approach A . A homeomorphic transformation in which angles between rays are preserved will be called *conformal*.

In general metric spaces angles defined in this manner obviously lack many important properties usually associated with angles and a further investigation of the types of spaces admitting these properties and of conditions for the existence of angles between rays is needed. We can, however, obtain one interesting result without restricting our space.

THEOREM. *Let X and Y be homeomorphic metric spaces and the transformation of X into Y be an unrestricted regular transformation whose scale function $\phi(u)$ has a derivative continuous at $u = 0$. Then the transformation is conformal.*

Proof. Let A be a fixed point, and B and C be variable points in X . Let $a = BC$, $b = AC$, and $c = AB$. Let the corresponding points and distances in Y be denoted by primed letters. To prove our theorem we must

* I. e., topological images of half-lines.

show that, if $\lim \cos BAC$ exists as B and C approach A , then $\lim \cos B'A'C'$ exists and is the same.

For any sequence of positions of B and C converging to A at least one of the following three cases for some sub-sequence will occur: (I) $a/b \rightarrow \infty$ or $a/c \rightarrow \infty$; (II) $a/b \rightarrow 0$ or $a/c \rightarrow 0$; (III) $a/b, a/c, b/a$, and c/a are bounded.

Case I. Let $a/b \rightarrow \infty$. By algebra

$$1 - \cos B'A'C' = (1/2) [1 + (a' - b')/c'] \cdot [1 + (a' - c')/b'].$$

By the mean value theorem we have $a' - b' = (a - b) \cdot \phi'(u')$ and $c' = c \cdot \phi'(u'')$, where $a < u' < b$ and $0 < u'' < c$. We can then write

$$(a' - b')/c' = [(a - b)/c] \cdot [\phi'(u')/\phi'(u'')].$$

As a, b , and c approach zero, $\lim [\phi'(u')/\phi'(u'')] = 1$. Since $a/b \rightarrow \infty$, it follows by the metric triangle inequality that $\lim [(a - b)/c] = 1$. Hence we know that $\lim [1 + (a' - b')/c'] = \lim [1 + (a - b)/c] = 2$. In consequence of the existence of $\lim (1 - \cos BAC)$ and of $\lim [1 + (a - b)/c]$, it follows that $\lim [(a - c)/b]$ exists. By an argument like that above, we then have $\lim [1 + (a' - c')/b'] = \lim [1 + (a - c)/b]$. Consequently, $\lim (1 - \cos B'A'C') = \lim (1 - \cos BAC)$. If $a/c \rightarrow \infty$, we proceed in exactly the same manner.

Case II. Let $a/b \rightarrow 0$ or $a/c \rightarrow 0$. The procedure is similar to that in Case I, with the exception that $\lim [(a - b)/c] = -1$, or $\lim [(a - c)/b] = -1$.

Case III. If $a/c, a/b, b/a$, and c/a are bounded, it is shown by the metric triangle inequality that b/c and c/b are also bounded. By simple algebra we find that

$$2(\cos B'A'C' - \cos BAC) = (b/c)[b'c/bc' - 1] + (c/b)[bc'/b'c - 1] + (a^2/bc)[1 - bca'^2/b'c'a^2].$$

Since $\phi'(0)$ is finite and not zero, each bracket converges to zero. On the other hand, the coefficients are all bounded. Hence the above equation gives $\lim \cos B'A'C' = \lim \cos BAC$.

Thus we have shown that every sequence of points B' and C' which are images of points B and C converging to A contains a sub-sequence for which $\lim \cos B'A'C' = \lim \cos BAC$. But it then follows from the theory of limits that the same equation is true for all sequences. Hence the theorem is proved.

8. *Smoothness of arcs.* Let A, B , and C be points of an oriented simple arc, with B and C on the right of A and the letters so assigned that $AB \leq AC$. If $\lim [(AB + BC - AC)/AB] = 0$ as B and C approach A , we say that the arc is *smooth on the right* at A . Similarly we define smoothness on the left.

If the arc is smooth on both sides of A and also for points B and C approaching A from opposite sides, with $AB \leq AC$, $\lim [(AB + AC - BC)/AB] = 0$, we say that the arc is *smooth about A* . In another paper by the writer* it is shown that for arcs in a Euclidean space these intrinsic conditions are necessary and sufficient for the existence of unilateral tangents and of tangents.

If the points are in the order A, B, C and both $\lim [(AB + BC - AC)/AB] = 0$ and $\lim [(AB + BC - AC)/BC] = 0$, we say that the arc is *strongly smooth on the right* at A . If the order is C, B, A , we have strong smoothness in the left. By application of the metric triangle inequality we can show that these conditions imply the condition for unilateral smoothness above. If an arc is strongly smooth on both sides of A and also smooth about A , we call it *strongly smooth about A* . It is readily seen from the plane cosine law that the conditions for unilateral strong smoothness are equivalent to requiring that angle $BAC \rightarrow 0$ and angle $ABC \rightarrow \pi$ as B and C approach A . If the arc lies in a Euclidean plane and has a tangent at each point B of a sub-arc AC , it then follows that strong smoothness on the right is equivalent to the requirement that the tangent turns in such a manner as to approach the right-hand tangent at A as B approaches A . Thus strong smoothness is a generalization for abstract spaces of the idea of a continuously turning tangent.

THEOREM. *Let X and Y be homeomorphic metric spaces and the transformation of X into Y be unrestricted and regular and its scale function $\phi(u)$ have a derivative continuous at $u = 0$. If a simple arc in X has smoothness of any kind at a point A , its image in Y has the same kind of smoothness at the corresponding point A' .*

Proof. Let us take the case that the given arc is smooth on the right at A ; the other cases are proved in exactly the same way. Let B and C be points of the arc at the right of A , $a = BC$, $b = AC$, $c = AB$, and $c \leq b$. Let the corresponding points and distances in Y be denoted by primed letters.

By definition $\lim [(c + a - b)/c] = 0$ as B and C approach A , whence $\lim [(a - b)/c] = -1$. By the mean value theorem $a' - b' = (a - b) \cdot \phi'(u')$ and $c' = c \cdot \phi'(u'')$, where $a < u' < b$ and $0 < u'' < c$. This gives $(a' - b')/c' = [(a - b)/c] \cdot [\phi'(u')/\phi'(u'')]$. By the continuity of $\phi'(u)$, $\lim [\phi'(u')/\phi'(u'')] = 1$. Hence $\lim [(a' - b')/c'] = \lim [(a - b)/c] = -1$, and $\lim [(c' + a' - b')/c'] = 0$, which was to be proved.

YALE UNIVERSITY,
NEW HAVEN, CONN.

* "On angles in certain metric spaces," *Bulletin of the American Mathematical Society*, vol. 38, pp. 580-588.

DISCONTINUOUS SOLUTIONS IN THE NON-PARAMETRIC PROBLEM OF MAYER IN THE CALCULUS OF VARIATIONS.*

By WILLIAM T. REID.

1. Introduction. By a discontinuous solution in a calculus of variations problem is meant an extremizing arc having one or more corners; that is, the derivatives of the functions defining the arc have one or more points of discontinuity. The treatment of discontinuous solutions has been largely limited to problems phrased in parametric form [see IV and V; the reader is referred to V for references to earlier literature on the subject].† In fact, Graves [V, p. 834] has stated that "it seems to be necessary to cast the problem in parametric form to get an effective extension of the Jacobi condition by means of the second variation."

The present paper treats discontinuous solutions for the general problem of Mayer in non-parametric form, and in terms of the characteristic numbers of a boundary value problem associated with the second variation there is given an effective extension of the Jacobi condition. Sufficient conditions are then established by the method used in the case of extremal arcs by Bliss and Hestenes and by Hestenes [VIII and IX]. The reader is also referred to the bibliographies at the end of these papers for references to earlier literature on the problem of Mayer. For the accessory boundary value problem in terms of which the analogue of the Jacobi condition is phrased the boundary conditions apply at more than two points, but it is significant that this problem may be transformed into one for which the boundary conditions apply at just two points, and of the type recently considered by the author [VII; also X, § 6]. In view of a recent paper by the author [X], it follows that an analogue of the Jacobi condition may equally well be phrased in a manner which is a direct generalization of that given by Bliss for the problem of Bolza, and also used by Hestenes for the general problem of Mayer [IX, p. 483].

In §§ 2 and 4 there are given conditions satisfied by a minimizing arc for the problem here considered, while § 3 considers the construction of families of extremaloids. The second variation, an analogue of Jacobi's condition in terms of an associated boundary value problem, together with the definition and determination of conjugate points, are discussed in §§ 5-8. In § 9 there is proved an auxiliary theorem which is fundamental in the

* Presented to the American Mathematical Society, June 23, 1933.

† Roman numerals in brackets refer to the bibliography at the end of this paper.

construction of a field of extremaloids as defined in § 10, and, as, used in § 11 to establish sufficient conditions for a strong relative minimum.

2. The first necessary condition for a minimum, and preliminary remarks. The general problem of Mayer as formulated by Bliss [I] is that of finding in a class of arcs

$$(2.1) \quad y_i = y_i(x) \quad (i = 1, \dots, n; x_1 \leq x \leq x_2)$$

satisfying a system of differential equations and end-conditions

$$(2.2) \quad \phi_\alpha[x, y, y'] = 0 \quad (\alpha = 1, \dots, m < n)$$

$$(2.3) \quad \psi_\mu[x_1, y(x_1), x_2, y(x_2)] = 0 \quad (\mu = 1, \dots, p \leq 2n + 1)$$

one which minimizes a function of the form

$$(2.4) \quad g[x_1, y(x_1), x_2, y(x_2)].$$

As is customary, we concentrate attention on a particular arc E_{12} with equations (2.1) and inquire what properties it must have if it is to be a minimizing arc. It will be assumed that the functions g, ψ_μ, ϕ_α and the functions $y_i(x)$ defining E_{12} satisfy the hypotheses used by Bliss and Hestenes [VIII, p. 306]. The reader is also referred to their paper for the definition of *admissible set* and *admissible arc*.

Suppose that E_{12} has r corners for values $x = t_\theta$ ($\theta = 1, \dots, r$), and $x_1 < t_1 < t_2 < \dots < t_r < x_2$. To make the notation simpler, let $t_0 = x_1$ and $t_{r+1} = x_2$. At times we shall be interested in the component arcs of E_{12} between corners. We therefore introduce functions $y_i^\gamma(x)$ ($\gamma = 1, \dots, r + 1$) defined by

$$(2.5) \quad y_i^\gamma(x) = y_i(x) \quad (t_{\gamma-1} \leq x \leq t_\gamma; \gamma = 1, \dots, r + 1),$$

and denote the sub-arcs of E_{12} thus obtained by E^{γ}_{12} . A one-parameter family of admissible arcs

$$y_i = y_i(x, b) \quad x_1(b) \leq x \leq x_2(b)$$

which contains E_{12} for $b = b_0$ and whose end-values satisfy the equations (2.3) may be thought of as a one-parameter family of arcs

$$(2.6) \quad y_i = y_i^\gamma(x, b) \quad (t_{\gamma-1}(b) \leq x \leq t_\gamma(b); \gamma = 1, \dots, r + 1)$$

which contains the set y_i^γ belonging to E^{γ}_{12} for $b = b_0$, which satisfies with the set $y_i^\gamma(x, b)$ the equations (2.2) on $t_{\gamma-1}(b) \leq x \leq t_\gamma(b)$, and is such that

$$(2.7) \quad \begin{array}{ll} \text{a)} & \psi_\mu[x_1(b), y^1(x_1(b), b), x_2(b), y^{r+1}(x_2(b), b)] = 0 \\ \text{b)} & \chi_i[\theta | y] \equiv y_i^\theta(t_\theta(b), b) - y_i^{\theta+1}(t_\theta(b), b) = 0 \quad (\theta = 1, \dots, r). \end{array}$$

The equations of variation of the differential equations (2.2) are

$$(2.8) \quad \Phi_\alpha[x, \eta, \eta'] \equiv \phi_{\alpha y'} \eta' + \phi_{\alpha y} \eta = 0 \quad (\alpha = 1, \dots, m),$$

where the coefficients $\phi_{\alpha y'}$, $\phi_{\alpha y}$ have as arguments the functions (2.1) defining E_{12} .

If we set

$$\xi_s = x_{sb}(b_0), \quad \tau_\theta = t_{\theta b}(b_0), \quad \eta_i = y_{ib}(x, b_0), \quad (s = 1, 2; \theta = 1, \dots, r),$$

then the equations of variation corresponding to (2.7) are

$$(2.9) \quad \begin{aligned} \text{a)} \quad \Psi_\mu[\xi, \eta] &\equiv (\psi_{\mu x_1} + \psi_{\mu y_{i1}} y'_{i1}) \xi_1 + (\psi_{\mu x_2} + \psi_{\mu y_{i2}} y'_{i2}) \xi_2 \\ &\quad + \psi_{\mu y_{i1}} \eta_i(x_1) + \psi_{\mu y_{i2}} \eta_i(x_2) = 0, \\ \text{b)} \quad X_i[\theta | \tau, \eta] &\equiv \{y'_i(t_{\theta^-}) - y'_i(t_{\theta^+})\} \tau_\theta + \eta_i(t_{\theta^-}) - \eta_i(t_{\theta^+}) = 0.* \end{aligned}$$

The corresponding variation of $g[x_1, y(x_1), x_2, y(x_2)]$ will be denoted by $G[\xi_1, \eta(x_1), \xi_2, \eta(x_2)]$.

A set of constants ξ_s, τ_θ and functions $\eta_i(x)$ will be called a *set of quasi-admissible variations* if on each interval $t_{\gamma-1} < x < t_\gamma$ the functions $\eta_i(x)$ coincide with functions $\eta_i^\gamma(x)$ which are continuous, consist of a finite number of pieces with continuous derivatives, and satisfy equations (2.8) on $t_{\gamma-1} \leq x \leq t_\gamma$ ($\gamma = 1, \dots, r+1$). A set of quasi-admissible variations will be called a *set of generalized admissible variations* if equations (2.9b) are satisfied. It follows readily that a set of generalized admissible variations is a set of admissible variations as defined by Bliss, if and only if $\tau_\theta = 0$ ($\theta = 1, \dots, r$).

If $\xi_{s\sigma}, \tau_{\theta\sigma}, \eta_{i\sigma}$ ($\sigma = 1, \dots, P+1 = p + rn + 1$) are $P+1$ sets of quasi-admissible variations, it follows by a method similar to that used by Bliss [III, p. 691] that there exists a family

$$(2.10) \quad \begin{aligned} y_i &= y_i(x, b_1, \dots, b_{P+1}), \\ t_{\gamma-1}(b_1, \dots, b_{P+1}) &\leq x \leq t_\gamma(b_1, \dots, b_{P+1}) \quad (\gamma = 1, \dots, r+1) \end{aligned}$$

satisfying equations (2.2), containing E_{12} for $b_\sigma = 0$, and having the sets $\xi_{s\sigma}, \tau_{\theta\sigma}, \eta_{i\sigma}$ as its variations along E_{12} with respect to the parameters b_σ . When the equations (2.10) are substituted in the functions g, ψ_μ, χ_i , these become functions of b_1, \dots, b_{P+1} . The following necessary condition may then be proved in the manner used by Bliss to establish the first necessary condition for the problems of Mayer [I, p. 311] and Lagrange [III, p. 683].

I. THE FIRST NECESSARY CONDITION. For every minimizing arc E_{12}

* Throughout this paper, the repetition of a subscript in an expression will indicate summation with respect to that subscript over its range of definition. The + and - signs attached to the argument of a function will be used to indicate the right-hand and left-hand limits of the function at the indicated value of the argument.

for the problem of Mayer with corners at $x = t_\theta$ ($\theta = 1, \dots, r$) there exists a function $F = \lambda_a(x)\phi_a$ such that between corners of E_{12} ,

$$(2.11) \quad (d/dx)F_{y'_i} - F_{y_i} = 0, \quad \phi_a = 0,$$

and such that

$$(2.12) \quad F_{y'_i}(x_2)\eta_i(x_2) - F_{y'_i}(x_1)\eta_i(x_1) + \lambda_0 G[\xi_1, \eta(x_1), \xi_2, \eta(x_2)] = 0$$

for every set $\xi_1, \eta_1(x_1), \xi_2, \eta(x_2)$ satisfying the equations $\Psi_\mu[\xi_1, \eta(x_1), \xi_2, \eta(x_2)] = 0$. The multipliers $\lambda_a(x)$ coincide on each segment $t_{\gamma-1} < x < t_\gamma$ with functions $\lambda_a^\gamma(x)$ which are continuous on $t_{\gamma-1} \leq x \leq t_\gamma$; furthermore, the functions $F_{y'_i}$, $F - y'_i F_{y'_i}$ are continuous at the corner points $x = t_\theta$ ($\theta = 1, \dots, r$).

This first necessary condition differs from that obtained in the usual fashion [IX, p. 480] in that it includes the additional condition that the function $F - y'_i F_{y'_i}$ is continuous at the corners of a minimizing arc.*

If E_{12} is an admissible arc which satisfies between corners the equations (2.11), and which is non-singular, that is, at each element (x, y, y', λ) of E_{12} the determinant

$$R = \begin{vmatrix} F_{y'_i y'_i} & \phi_{a y'_i} \\ \phi_{\beta y'_i} & 0 \end{vmatrix}$$

is different from zero, then the functions $y_i(x)$ and $\lambda_a(x)$ are of classes C'' and C' at least between corners. In this case E_{12} consists of a finite number of extremal arcs and is called a broken extremal. A broken extremal such that the functions $F_{y'_i}$ and $F - y'_i F_{y'_i}$ are continuous at its corners will be called an *extremaloid*.

The admissible arc E_{12} is normal if there exist p sets of admissible variations ξ_{sv}, η_{iv} ($s = 1, 2; v = 1, \dots, p$) such that

$$(2.13) \quad |\Psi_\mu[\xi_v, \eta_v]| \neq 0.$$

E_{12} is seen to be normal if and only if there exist P sets of quasi-admissible variations $\xi_{i\pi}, \tau_{\theta\pi}, \eta_{i\pi}$ ($\pi = 1, \dots, P$) such that

$$\begin{vmatrix} \Psi_\mu[\xi_\pi, \eta_\pi] \\ X_i[1 | \tau_\pi, \eta_\pi] \\ \cdot \\ X_i[r | \tau_\pi, \eta_\pi] \end{vmatrix} \neq 0.^\dagger$$

* The continuity of $F - y'_i F_{y'_i}$ has been derived in another manner by Bliss and Hestenes, who have also discussed the dependence of the continuity of this function upon the continuity of $F_{y'_i}$. For abstract of their paper, see *Bulletin of the American Mathematical Society*, vol. 39 (1933), p. 341.

† One such set is obtained as follows: let $\tau_{\theta\pi} = 0$ ($\pi = 1, \dots, P; \theta = 1, \dots, r$),

The admissible arc E_{12} is normal on the sub-interval $x'x''$ if there exist $2n-1$ sets of admissible variations $\xi_{sk}, \eta_{ik}(x)$ ($k=1, \dots, 2n-1$) such that the matrix

$$\begin{vmatrix} \eta_{ik}(x') \\ \eta_{ik}(x'') \end{vmatrix}$$

has rank $2n-1$.

The method used by Bliss to prove the corresponding result for the problem of Lagrange [III, p. 695], gives the following result.

LEMMA 2.1. If a set of generalized admissible variations $\xi_s, \tau_\theta, \eta_i(x)$ for a normal admissible arc E_{12} satisfy the equations $\Psi_\mu[\xi, \eta] = 0$, then there exists a one-parameter family of admissible arcs

$$(2.14) \quad y_i = y_i^\gamma(x, b) \quad t_{\gamma-1}(b) \leq x \leq t_\gamma(b), \quad (\gamma = 1, \dots, r+1)$$

satisfying the equations (2.7), containing E_{12} for $b = b_0$, and having the set $\xi_s, \tau_\theta, \eta_i(x)$ as its variations along E_{12} .

We have also:

LEMMA 2.2. If E_{12} is an extremaloid, then for every set of generalized admissible variations $\xi_s, \tau_\theta, \eta_i(x)$ along E_{12} the functions $\eta_i(x)$ satisfy the relation

$$(2.15) \quad F_{y', \eta_i} \bigg|_{x=x'}^{x=x''} = 0$$

for every set of values x' and x'' on x_1x_2 .

It then follows readily that at each corner $x = t_\theta$ of E_{12} the functions $\eta_i(x)$ belonging to a set of generalized admissible variations $\xi_s, \tau_\theta, \eta_i(x)$ satisfy the equation

$$(2.16) \quad F_{y', \eta_i}(t_{\theta-})\eta_i(t_{\theta-}) - F_{y', \eta_i}(t_{\theta+})\eta_i(t_{\theta+}) = 0.$$

3. Construction of families of extremaloids. Embedding theorems for extremaloids will be obtained by starting with an extremal sub-arc of the given extremaloid and showing how to proceed past a corner. We shall assume that the given extremaloids E_{12} has corners at $x = t_\theta$ ($\theta = 1, \dots, r$) and is non-singular. Along an extremal sub-arc E_{12}^γ of E_{12} whose equations are of the form (2.5), the equations

and choose ξ_{sv}, η_{iv} ($v = 1, \dots, p$) as a set of admissible variations satisfying (2.13); finally choose quasi-admissible variations $\eta_{i, p+(\theta-1)n+j}(x)$ as continuous on $x_1 \leq x \leq t_\theta$, satisfying the conditions $\eta_{i, p+(\theta-1)n+j}(t_\theta) = \delta_{ij}$, and as identically zero on $t_\theta \leq x \leq x_2$ ($\theta = 1, \dots, r$).

$$(3.1) \quad F_{y'_i}(x, y, y', \lambda) = z_i, \quad \phi_a(x, y, y') = 0$$

may be solved for the variables y'_i, λ_a in a neighborhood of the values (x, y, z) on E^{γ}_{12} . The solution of (3.1) has the form

$$(3.2) \quad y'_i = P_i^{\gamma}(x, y, z), \lambda_a = \Lambda_a^{\gamma}(x, y, z),$$

and has continuous partial derivatives of the first two orders since the first members of the equations (3.1) have such derivatives. The system (2.11) is now equivalent to the system

$$(3.3) \quad dy_i/dx = P_i^{\gamma}(x, y, z), \quad dz_i/dx = F_{y_i}[x, y, P^{\gamma}(x, y, z), \Lambda^{\gamma}(x, y, z)].$$

It then follows that through each element (x_0, y_0, z_0) in a neighborhood of the set of values (x, y, z) on E^{γ}_{12} there passes a unique solution

$$(3.4) \quad y_i = Y_i^{\gamma}(x, x_0, y_0, z_0), \quad z_i = Z_i^{\gamma}(x, x_0, y_0, z_0)$$

of equations (3.3) for which the functions $Y_i^{\gamma}, Y_{ix}^{\gamma}, Z_i^{\gamma}, Z_{ix}^{\gamma}$ have continuous partial derivatives of the first two orders since the second members of (3.3) have such derivatives [See III, § 6 and VIII, § 3].

For (x_0, y_0, z_0) in a suitably restricted neighborhood of E^1_{12} we then have a family of extremal arcs of the form (3.4) for $\gamma = 1$ embedding E^1_{12} . Now along this family the corner equations

$$(3.5) \quad \begin{aligned} &F_{y'_i}[T, Y^1(T), Y^1_x(T), \Lambda^1(T)] - F_{y'_i}[T, Y^1(T), p, \lambda] = 0, \\ &F[T, Y^1(T), Y^1_x(T), \Lambda^1(T)] - Y^1_{ix}(T)F_{y'_i}[T, Y^1(T), Y^1_x(T), \Lambda^1(T)] \\ &\quad - F[T, Y^1(T), p, \lambda] + p_i F_{y'_i}[T, Y^1(T), p, \lambda] = 0, \\ &\phi_a[T, Y^1(T), p] = 0, \end{aligned}$$

where for brevity the arguments (x_0, y_0, z_0) are omitted in writing, have initial solutions $T = t_1$, $p_i = y'_i(t_1^+)$, $\lambda_a = \lambda_a(t_1^+)$, (x_0, y_0, z_0) equal to an arbitrary element x_{00}, y_{00}, z_{00} on E^1_{12} . Since E_{12} is non-singular, the functional determinant of the equations (3.5) with respect to the variables p_i, λ_a, T is seen to be equal to a non-zero multiple of the value of Ω_0 at $x = t_1$, where

$$(3.6) \quad \Omega_0(x) = F_{x^+} + y'_i F_{y_i^+} - F_{x^-} - y'_i F_{y_i^-},$$

and the superscripts $+$ and $-$ denote, respectively, the right-hand and left-hand limits at a given value of x .

Consequently, if we make the additional assumption that $\Omega_0(t_1) \neq 0$, the equations (3.5) have unique solutions $p_i = p_i(x_0, y_0, z_0)$, $\lambda_a = \lambda_a(x_0, y_0, z_0)$, $T = T_1(x_0, y_0, z_0)$ having $x_0, y_0, z_0, p_i, \lambda_a, T$ in suitably restricted neighborhoods of $x_{00}, y_{00}, z_{00}, y'_i(t_1^+), \lambda_a(t_1^+), t_1$, respectively, and these solutions are of class C' at least. Application of the embedding theorem to the extremal sub-arc E^2_{12} now shows that through each element

$$[T_1(x_0, y_0, z_0), Y^1(T_1(x_0, y_0, z_0), x_0, y_0, z_0), p_i(x_0, y_0, z_0), \lambda_a(x_0, y_0, z_0)]$$

there passes a unique extremal arc in a neighborhood of E_{12}^2 , provided (x_0, y_0, z_0) is sufficiently near (x_{00}, y_{00}, z_{00}) . If Ω_0 is assumed to be different from zero at each corner of E_{12} we may proceed past these corners successively and obtain a family of extremaloids

$$(3.7) \quad y_i = Y_i(x, x_0, y_0, z_0), \quad z_i = Z_i(x, x_0, y_0, z_0) \quad (x_1 \leq x \leq x_2)$$

defined for (x, x_0, y_0, z_0) in a neighborhood of the values corresponding to E_{12} and containing E_{12} for $(x_1, x_2, x_0, y_0, z_0) = (x_{10}, x_{20}, x_{00}, y_{00}, z_{00})$. The functions Y_i are continuous in all their arguments, and $Y_i, Y_{ix}, Y_{ixx}, \lambda_a, \lambda_{ax}$ are of class C' in all their arguments except at the corners; furthermore, the corner manifolds

$$(3.8) \quad x = T_\theta(x_0, y_0, z_0), y_i = Y_i[T_\theta(x_0, y_0, z_0), x_0, y_0, z_0] \quad (\theta = 1, \dots, r)$$

are of class C' .

In the above proof of the existence of the family (3.7) we have assumed that the element (x_{00}, y_{00}, z_{00}) was on E_{12}^1 . If this initial element occurred on some other extremal sub-arc of E_{12} it is clear that a family of the form (3.7) would be obtained by extending the family past the corners of the adjoining sub-arcs successively.

Since each curve (3.7) has an initial set at $x = x_{10}$ we lose none of them if we replace x_0 by the fixed value x_{10} . Furthermore, on account of homogeneity relations satisfied by the functions of the set (3.7), the family is seen to depend essentially upon only $2n - 1$ constants [VIII, p. 310]. If these constants are suitably chosen and denoted by c_1, \dots, c_{2n-1} , the equations of the family take the form

$$(3.9) \quad y_i = y_i(x, c), \quad z_i = z_i(x, c) \quad (x_1 \leq x \leq x_2)$$

and contains E_{12} for special values $(x_1, x_2, c) = (x_{10}, x_{20}, c_0)$. In a neighborhood of the values (x, c) defining E_{12} the functions

$$y_i, y_{ix}, y_{ixx}, z_i, z_{ix}, \lambda_a = \Lambda_a[x, y(x, c), z(x, c)], \lambda_{ax}$$

are of class C' in all the arguments except along the corner curves of the family, and for the special values (x_{10}, c_0) the determinant

$$(3.10) \quad \begin{vmatrix} y_{ic_s} & 0 \\ z_{ic_s} & z_i \end{vmatrix}$$

is different from zero. Finally, the equations of the corner manifolds of the family may be written in the form

$$\mathfrak{M}_\theta(c) : x = t_\theta(c), \quad y_i = y_i(t_\theta(c), c) \quad (\theta = 1, \dots, r)$$

and the defining functions are of class C' in a neighborhood of $c = c_0$.

4. Further conditions satisfied by a minimizing arc. The Weierstrass necessary condition states that at each element (x, y, y', λ) of a normal minimizing arc E_{12} the inequality

$$(4.1) \quad \mathcal{E}[x, y, y', Y', \lambda] \\ \equiv F(x, y, Y', \lambda) - F(x, y, y', \lambda) - (Y' - y')F_{y'}(x, y, y', \lambda) \geq 0$$

must be satisfied for every admissible set $(x, y, Y') \neq (x, y, y')$.^{*} The necessary condition of Clebsch states that at each element (x, y, y', λ) of a normal minimizing arc the inequality

$$F_{y'y'}(x, y, y', \lambda)\Pi_i\Pi_i \geq 0$$

must be satisfied by every set $(\Pi_i) \neq (0)$ which satisfies the equations $\phi_{ay'}\Pi_i = 0$. The conditions of Weierstrass and Clebsch will be denoted by II and III, respectively, while II' and III' will be used to denote the conditions strengthened to exclude the equality sign.

At a corner $x = t_\theta$ the corner conditions of the first necessary condition imply

$$(4.2) \quad \mathcal{E}[t_\theta, y(t_\theta), y'(t_\theta^-), y'(t_\theta^+), \lambda(t_\theta^-)] \\ = 0 = \mathcal{E}[t_\theta, y(t_\theta), y'(t_\theta^+), y'(t_\theta^-), \lambda(t_\theta^+)].$$

Let $P_i^\theta(x)$ be a set of continuous functions such that $P_i^\theta(t_\theta) = y'_i(t_\theta^+)$ and which satisfies the equations $\phi_a[x, y^\theta(x), P^\theta(x)] = 0$ in a neighborhood of $x = t_\theta$. By the use of the corner conditions we obtain that the function $\mathcal{E}[x, y(x), y'(x), P^\theta(x), \lambda(x)]$ has at $x = t_\theta$ a left-handed derivative equal in value to $\Omega_0(t_\theta)$.[†] Similarly, if $Q_i^\theta(x)$ is a set of continuous functions satisfying $Q_i^\theta(t_\theta) = y'_i(t_\theta^-)$ and $\phi_a[x, y^{\theta+1}(x), Q^\theta(x)] = 0$ in a neighborhood of $x = t_\theta$, then $\mathcal{E}[x, y(x), y'(x), Q^\theta(x), \lambda(x)]$ has a right-handed derivative at $x = t_\theta$ equal to $-\Omega_0(t)$. We have, therefore, the following results:

If E_{12} is an extremaloid along which $\mathcal{E} \geq 0$, then $\Omega_0 \leq 0$ at the corners of E_{12} .

If E_{12} is a minimizing arc for the problem of Mayer, then $\Omega_0 \leq 0$ at the corners of E_{12} .

We have also the following property:

If E_{12} is an extremal arc and at an element (x, y, y', λ) of E_{12} the \mathcal{E} function is non-negative, but vanishes for an admissible set (x, y, Y')

^{*} See Graves, "On the Weierstrass condition for the problem of Bolza in the calculus of variations," *Annals of Mathematics*, vol. 33 (1932), pp. 261-274.

[†] The corresponding result for the Lagrange problem in parametric form has been established in the above manner by Hefner [VI].

$\neq (x, y, y')$; then there exist multipliers $\bar{\lambda}_a$ such that the set $(x, y, Y', \bar{\lambda})$ satisfies the corner conditions with the set (x, y, y', λ) .

The function $\mathcal{E}[x, y, y', p, \lambda]$ is non-negative for all sets p_i such that (x, y, p) is in \mathfrak{R} , $\phi_a[x, y, p] = 0$, and $\|\phi_{ay'}(x, y, p)\|$ is of rank m . Therefore, since $p_i = Y'_i$ is a set such that

$$(4.3) \quad \mathcal{E}[x, y, y', Y', \lambda] = 0,$$

it follows that there exist constants d_a satisfying the equations

$$(\lambda_a + d_a)\phi_{ay'}(x, y, Y') - \lambda_a\phi_{ay'}(x, y, y') = 0.$$

If $\bar{\lambda}_a = \lambda_a + d_a$, the set $(x, y, Y', \bar{\lambda})$ satisfies the conditions

$$F_{y'_i}(x, y, Y', \bar{\lambda}) = F_{y'_i}(x, y, y', \lambda),$$

and from (4.3) it then follows that the other corner conditions are also satisfied.

In view of equations (4.2) it is seen that along an extremaloid which is not an extremal arc condition II' cannot be satisfied. As a matter of notation, however, an arc of the form (2.1) will be said to satisfy condition II', if at each element (x, y, y', λ) on the arc we have

$$\mathcal{E}[x, y, y', Y', \lambda] > 0$$

for every admissible set (x, y, Y') which is distinct from the sets $[x, y, y'(x-)]$ and $[x, y, y'(x+)]$.

5. The second variation. Consider a normal non-singular extremaloid E_{12} with end-values satisfying the equations $\psi_\mu = 0$ and which has corners at $x = t_\theta$ ($\theta = 1, \dots, r$). Let $\xi_s, \tau_\theta, \eta_i(x)$ be a set of generalized admissible variations satisfying the equations $\Psi_\mu[\xi, \eta] = 0$. By Lemma 2.1 there is a one-parameter family of admissible arcs (2.14) containing E_{12} for $b = b_0$, satisfying $\psi_\mu = 0$, and having the set $\xi_s, \tau_\theta, \eta_i(x)$ as its variations along E_{12} . If the equations of this family are substituted in the functions $g, \psi_\mu, \chi_i, \phi_a$ we obtain*

$$(5.1) \quad g_2[\xi, \tau, \eta] = \frac{d^2 g}{db^2} \Big|_{b=b_0} = 2Q[\xi_1, \eta(x_1); \xi_2, \eta(x_2)] \\ + \sum_{\theta=1}^r 2H[\theta | \tau, \eta(t_\theta^-), \eta(t_\theta^+)] + \int_{x_1}^{x_2} 2\omega(x, \eta, \eta') dx$$

where

$$2\omega = F_{y'y'}\eta'\eta' + 2F_{y'y}\eta'\eta + F_{yy}\eta\eta,$$

$$H[\theta | \tau_\theta, \eta(t_\theta^-), \eta(t_\theta^+)] = [\tau_\theta^2 \{F_x + F_{y_i}y'_i\} + 2\tau_\theta F_{y_i}\eta_i] \Big|_{x=t_\theta^-}^{x=t_\theta^+},$$

* The reader is referred to [II] for the corresponding calculation of the second variation along an extremal arc.

and Q is a quadratic form in its arguments whose explicit value will not be written.

6. The accessory boundary value problem and an analogue of Jacobi's condition. We shall now consider the problem of minimizing the expression $g_2[\xi, \tau, \eta]$ in the class of generalized admissible variations which satisfy the equations $\Psi_\mu[\xi, \eta] = 0$, and are normed such that

$$(6.1) \quad \xi_s \xi_s + \tau_\theta \tau_\theta + \int_{x_1}^{x_2} \eta_i(x) \eta_i(x) dx = 1.$$

The formulation of this problem is simplified by the introduction of new functions $u_k(x)$ [$k = 1, \dots, n(r+1)$] by the equations

$$(6.2) \quad u_{(\gamma-1)n+i}(x) = \eta_i(t_{\gamma-1} + x[t_\gamma - t_{\gamma-1}]) \quad (0 \leq x \leq 1; \gamma = 1, \dots, r+1).$$

The limiting values of the functions $\eta_i(x)$ at the points $x = t_\theta$ are then expressible as end-values of the functions $u_k(x)$ at $x = 0$ and $x = 1$. The integrals in (5.1) and (6.1) are seen to become integrals on $0 \leq x \leq 1$ of quadratic forms in the variables $u_k(x)$ and their derivatives, and the set of m equations (2.8) reduces to a set of $m(r+1)$ linear differential equations in the new variables. Finally, let $r+2$ additional functions be defined by the equations

$$\begin{aligned} u_{rn+n+s}(0) &= \xi_s, \quad u_{rn+n+2+\theta}(0) = \tau_\theta & (s = 1, 2; \theta = 1, \dots, r), \\ u'_{rn+n+s} &= 0 = u'_{rn+n+2+\theta} & (0 \leq x \leq 1). \end{aligned}$$

Our accessory minimum problem described above then reduces to the problem of finding in a class of arcs

$$(6.3) \quad u_\nu = u_\nu(x) \quad (0 \leq x \leq 1; \nu = 1, \dots, N = rn + r + n + 2),$$

which satisfy a set of $mr + m + r + 2$ ordinary linear homogeneous differential equations of the first order, $rn + p$ linear homogeneous equations in the end-values of these functions at $x = 0$ and $x = 1$, and the condition

$$(6.4) \quad \sum_{v=0}^{r+2} \{u_{rn+n+v}(0)\}^2 + \int_0^1 \sum_{\gamma=1}^{r+1} [\sum_{i=1}^n u_{(\gamma-1)n+i}^2(x)] \cdot [t_\gamma - t_{\gamma-1}] dx = 1,$$

one which minimizes an expression of the form

$$(6.5) \quad I_2[u] = 2\mathcal{Q}[u(0), u(1)] + \int_0^1 2w(x, u, u') dx.$$

In this last expression \mathcal{Q} and w are quadratic forms in $u_\nu(0), u_\nu(1)$ and u_ν, u'_ν respectively.

The boundary value problem consisting of the Euler-Lagrange equations and transversality conditions for this reduced form of the accessory minimizing problem is of the form recently considered by the author [VII, and

X, § 6]. If E_{12} is a normal extremaloid which satisfies condition III', then the conditions of Theorem 6.1 of [X] are satisfied and there exist infinitely many real characteristic numbers $\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots$ of the boundary value problem. Furthermore, each characteristic number is determined by a corresponding minimizing property. In particular, λ_1 is the minimum of $I_2[u]$ in the class of arcs (6.3) which satisfy the differential equations, end-conditions, and condition (6.4).

It is more convenient for our use, however, to write the differential equations and boundary conditions of this accessory boundary value problem in terms of the original set of variations $\xi_s, \tau_\theta, \eta_i(x)$ defined on x_1x_2 . It is also advantageous to give the differential equations in canonical form. As is customary, let

$$\Omega(x, \eta, \eta', \mu) = \omega(x, \eta, \eta') + \mu_a \Phi_a(x, \eta, \eta').$$

Along a non-singular extremaloid E_{12} the equations

$$(6.7) \quad \xi_i = \Omega_{\eta'_i}(x, \eta, \eta', \mu), \quad \Phi_a(x, \eta, \eta') = 0$$

have unique solutions

$$(6.8) \quad \eta'_i = H_i[x, \eta, \xi], \quad \mu_a = M_a[x, \eta, \xi]$$

which are linear in the variables η_i, ξ_i . It then follows that the characteristic numbers of the accessory boundary value problem are the values of λ corresponding to which there exist sets $\xi_s, \tau_\theta, \eta_i(x), \xi_i(x)$ which satisfy the following conditions:

- (a) the set $\xi_s, \tau_\theta, \eta_i(x)$ is a set of generalized admissible variations;
- (b) the functions $\eta_i(x), \xi_i(x)$ coincide on each segment $t_{\gamma-1} < x < t_\gamma$ with functions which are of class C' on $t_{\gamma-1} \leq x \leq t_\gamma$;
- (c) the functions η_i, ξ_i are not all identically zero on x_1x_2 , and satisfy on each segment $t_{\gamma-1} < x < t_\gamma$ the equations

$$(6.9) \quad \eta'_i = H_i[x, \eta, \xi], \quad \xi'_i = \Omega_{\eta'_i}(x, \eta, H[x, \eta, \xi], M[x, \eta, \xi]) - \lambda \eta_i = 0, \quad \Phi_a = 0;$$

- (d) there exist constants

$$d_\mu, d_{\theta|i} \quad (\mu = 1, \dots, p; \theta = 1, \dots, r; i = 1, \dots, n)$$

which satisfy the equations

$$(6.10) \quad \begin{aligned} \Psi_\mu[\xi, \eta] &= 0 & (\mu = 1, \dots, p), \\ X_i[\theta | \tau, \eta] &= 0 & (i = 1, \dots, n; \theta = 1, \dots, r), \\ Q_{\eta_{is}} + d_\mu \Psi_{\mu \eta_{is}} + (-1)^s \xi_i(x_s) &= 0, \\ Q_{\xi_s} + d_\mu \Psi_{\mu \xi_s} - \lambda \xi_s &= 0 & (s = 1, 2), \\ \tau_\theta F_{y_i}(t_\theta^-) + d_{\theta|i} + \xi_i(t_\theta^-) &= 0, \\ \tau_\theta F_{y_i}(t_\theta^+) + d_{\theta|i} + \xi_i(t_\theta^+) &= 0, \\ H_{\tau_\theta}[\theta | \tau, \eta] + \{y'_i(t_\theta^-) - y'_i(t_\theta^+)\} d_{\theta|i} - \lambda \tau_\theta &= 0. \end{aligned}$$

It is to be remarked that if $\xi_s, \tau_\theta, \eta_i, \xi_i$ is a solution of the system (6.9), (6.10), then the multipliers $\mu_a(x)$ defined by (6.8), together with the functions $\eta'_i(x)$, coincide on $t_{\gamma-1} < x < t_\gamma$ with functions which are of class C' on $t_{\gamma-1} \leq x \leq t_\gamma$ ($\gamma = 1, \dots, r+1$).

We have the following important theorem.

THEOREM 6.1. *Let E_{12} be a normal admissible arc whose end-values satisfy the equations $\Psi_\mu = 0$, and which satisfies conditions I and III'. Then a necessary and sufficient condition that $g_2[\xi, \tau, \eta] \geq 0$ for all sets of generalized admissible variations $\xi_s, \tau_\theta, \eta_i(x)$ which satisfy the equations $\Psi_\mu = 0$ is that the least characteristic number λ_1 of the corresponding boundary value problem (6.9), (6.10) be non-negative.*

As a corollary to this theorem we have the following result:

IV. THE CONDITION OF MAYER. *For a normal non-singular minimizing arc E_{12} the least characteristic number λ_1 of the corresponding accessory boundary value problem (6.9), (6.10) must satisfy the condition $\lambda_1 \geq 0$.*

As is customary, IV' will be used to denote condition IV with the equality sign excluded.

7. Secondary extremaloids. The differential equations (6.9) reduce for $\lambda = 0$ to the set

$$(7.1) \quad \eta'_i = H_i[x, \eta, \xi], \quad \xi'_i = \Omega_{\eta_i}(x, \eta, H[x, \eta, \xi], M[x, \eta, \xi]), \quad \Phi_a(x, \eta, \eta') = 0.$$

This set is called the *accessory system of differential equations*. We shall understand by a *secondary extremaloid* a set of functions $\eta_i(x)$, $\xi_i(x)$ and constants τ_θ satisfying condition (b) of § 6, which is such that equations (7.1) are satisfied on each segment $t_{\gamma-1} < x < t_\gamma$, and which satisfies with associated constants $d_{\theta|i}$ the corner equations

$$(7.2) \quad \begin{aligned} X_i[\theta | \tau, \eta] &= 0, \\ X_{n+i}[\theta | \tau, \eta, d] &\equiv \tau_\theta F_{y_i}(t_{\theta^-}) + d_{\theta|i} + \xi_i(t_{\theta^-}) = 0, \\ X_{2n+i}[\theta | \tau, \eta, d] &\equiv \tau_\theta F_{y_i}(t_{\theta^+}) + d_{\theta|i} + \xi_i(t_{\theta^+}) = 0, \\ X_{3n+i}[\theta | \tau, \eta, d] &\equiv H_{\tau_\theta}[\theta | \tau, \eta] + \{y'_i(t_{\theta^-}) - y'_i(t_{\theta^+})\} d_{\theta|i} = 0 \\ &\quad (\theta = 1, \dots, r). \end{aligned}$$

Equations (7.2) are seen to be those of the set (6.10) which apply at the points $x = t_\theta$ with $\lambda = 0$.

The following lemma is of significance:

LEMMA 7.1. *If E_{12} is a normal non-singular extremaloid having $\Omega_0 \neq 0$ at its corners, then corresponding to a secondary extremal arc η_i, ξ_i defined on one of the intervals $t_{\gamma-1} \leq x \leq t_\gamma$ there is a unique secondary extremaloid on $x_1 \leq x \leq x_2$ whose defining functions coincide with η_i, ξ_i on their interval of definition.*

In the $3n + 1$ equations of (7.2) corresponding to a given value of θ the determinant of the coefficients of the quantities $\tau_\theta, d_{\theta|1}, \eta_i(t_\theta^-), \xi_i(t_\theta^-)$, or of the quantities $\tau_\theta, d_{\theta|1}, \eta_i(t_\theta^+), \xi_i(t_\theta^+)$, is seen to be a non-zero multiple of $\Omega_0(t_\theta)$. Lemma 7.1 then follows readily from the hypothesis $\Omega_0(t_\theta) \neq 0$ and the existence theorem for the system (7.1).

In particular, it is to be noted that $\eta_i \equiv 0, \xi_i \equiv z_i(x)$ is a secondary extremaloid for which $\tau_\theta = 0$ ($\theta = 1, \dots, r$).

8. The determination of conjugate points. As before, we consider a normal non-singular extremaloid E_{12} with corners at $x = t_\theta$ ($\theta = 1, \dots, r$).

DEFINITION OF CONJUGATE POINT. A value x_3 is said to define a point 3 conjugate to the point 1 on E_{12} if there exists a secondary extremaloid $\tau_\theta = \sigma_\theta, \eta_i = u_i(x), \xi_i = v_i(x)$ whose functions $u_i(x)$ are not all identically zero on x_1x_3 , and satisfy with constants c, d the conditions

$$(8.1) \quad u_i(x_1) = 0 = cu_i(x_3^-) + du_i(x_3^+) \quad (cd \geq 0, c + d \neq 0).$$

If x_3 does not define a corner of E_{12} the conditions (8.1) become $u_i(x_1) = 0 = u_i(x_3)$.

The following lemma is a consequence of equations (7.2) and (3.6), and the proof will be omitted. The results will be used in establishing certain properties of conjugate points.

LEMMA 8.1. Let $\tau_\theta, \eta_i(x), \xi_i(x)$ be a secondary extremaloid which at a particular corner point $x = t_\theta$ of E_{12} satisfies with constants c and d the relation

$$(8.2) \quad c\eta_i(t_\theta^-) + d\eta_i(t_\theta^+) = 0 \quad (c + d \neq 0).$$

Then

$$(8.3) \quad 2H[\theta | d\tau_\theta/(c + d), \eta(t_\theta^-), 0] + \eta_i(t_\theta^-)\xi_i(t_\theta^-) = cd\tau_\theta^2\Omega_0(t_\theta)/(c + d)^2,$$

$$(8.4) \quad 2H[\theta | c\tau_\theta/(c + d), 0, \eta(t_\theta^+)] - \eta_i(t_\theta^+)\xi_i(t_\theta^+) = cd\tau_\theta^2\Omega_0(t_\theta)/(c + d)^2.$$

A generalization of the result obtained by Bliss and Hestenes [VIII, p. 315] for extremal arcs in the special Mayer problem for which $p = 2n + 1$ is given by the following necessary condition.

IV₀. Let E_{12} be a normal non-singular extremaloid, normal on every sub-interval x_3x_2 of E_{12} , and having $\Omega_0 \neq 0$ at each of its corners. If E_{12} is a minimizing arc for the problem of Mayer here proposed, then between 1 and 2 on E_{12} there can be no points 3 conjugate to 1.

If there were a secondary extremaloid σ_θ, u_i, v_i whose functions $u_i(x)$ satisfied conditions (8.1) but were not all identically zero on x_1x_3 , then let

$$(8.5) \quad \eta_i(x) = u_i(x) \quad (x_1 \leq x \leq x_3), \quad \eta_i(x) \equiv 0 \quad (x_3 < x \leq x_3).$$

In case either $u_i(x_3^-) = 0$ or $u_i(x_3^+) = 0$ ($i = 1, \dots, n$), corresponding constants τ_θ would be defined as follows: if $u_i(x_3^-) = 0$ ($i = 1, \dots, n$) then $\tau_\theta = \sigma_\theta$ if $t_\theta < x_3$ and $\tau_\theta = 0$ if $t_\theta \geq x_3$; if the values $u_i(x_3^-)$ were not all zero but $u_i(x_3^+) = 0$ ($i = 1, \dots, n$), then $\tau_\theta = \sigma_\theta$ if $t_\theta \leq x_3$ and $\tau_\theta = 0$ if $t_\theta > x_3$. These values, together with the functions η_i of (8.5) and the constants $\xi_s = 0$, would be such that $g_2[\xi, \tau, \eta] = 0$, and would be therefore a minimizing set for the accessory minimum problem of § 6. There would then exist multipliers $\mu_\alpha(x)$ and functions $\zeta_i(x)$ defined by (6.7) which were of class C' and satisfied with the functions (8.5) equations (7.1) on every sub-interval of x_1x_2 which did not contain any of the points x_3, t_θ ($\theta = 1, \dots, r$). Furthermore, the functions $\zeta_i(x)$ would be continuous between corners of E_{12} and equations (7.2) would be satisfied by suitable constants $d_{\theta|i}$. In view of the normality of E_{12} on x_3x_2 , there would then exist a constant k such that $\zeta_i(x) = kz_i(x)$ on x_3x_2 , and as a consequence of the continuity of $\zeta_i(x)$ between corners of E_{12} and Lemma 7.1 it would follow that $u_i(x) \equiv 0$ on x_1x_3 , which is a contradiction.

There still remains the possibility of x_3 defining a corner point t_{θ^*} of E_{12} and the functions of neither of the sets $u_i(x_3^-)$, $u_i(x_3^+)$ being all zero. In this case it would follow from the equations $X_i[\theta^* | \sigma, u] = 0$ that $\sigma_{\theta^*} \neq 0$. With the functions $\eta_i(x)$ of (8.5) we would then associate constants ξ_s, τ_θ as follows:

$$\xi_s = 0 \quad (s = 1, 2); \quad \tau_\theta = \sigma_\theta \text{ if } \theta < \theta^*, \quad \tau_\theta = 0 \text{ if } \theta > \theta^*, \quad \tau_{\theta^*} = d\sigma_{\theta^*}/(c + d).$$

It would then follow from Lemma 8.1 that

$$g_2[\xi, \tau, \eta] = cd\sigma_{\theta^*}^2\Omega_0(t_{\theta^*})/(c + d)^2 < 0.$$

Since this is impossible if E_{12} is a minimizing arc, the necessary condition IV_0 is therefore established.

The notation IV'_0 will be used to denote the condition that there exist no value x on $x_1 < x \leq x_2$ which defines a point conjugate to the point 1 on E_{12} . It follows readily that condition IV' implies IV'_0 .

If $\bar{\eta}_i, \bar{\xi}_i$ and \bar{u}_i, \bar{v}_i are a pair of secondary extremal arcs on a segment $t_{\gamma-1} < x < t_\gamma$, then the expression $\bar{\eta}_i\bar{v}_i - \bar{\xi}_i\bar{u}_i$ is equal to a constant on this segment [III, p. 738]; in particular, if this constant is zero the secondary extremal arcs are said to be conjugate. Let $\tau_\theta, \eta_i, \xi_i$ and σ_θ, u_i, v_i denote the secondary extremaloids on x_1x_2 whose functions η_i, ξ_i and u_i, v_i coincide with $\bar{\eta}_i, \bar{\xi}_i$ and \bar{u}_i, \bar{v}_i , respectively, on their interval of definition. From the conditions (7.2) satisfied by these secondary extremaloids it follows that the function $\eta_i v_i - \xi_i u_i$ is constant on x_1x_2 . Hence, if on any segment $t_{\gamma-1} < x < t_\gamma$ the component secondary extremal arcs of two secondary extremaloids are conjugate, then on each segment $t_{\gamma-1} < x < t_\gamma$ ($\gamma = 1, \dots, r + 1$)

the component secondary extremal arcs are conjugate; two such secondary extremaloids are said to be conjugate.

The following lemma will be of use in the proof of the succeeding theorem.

LEMMA 8.2. *Let $\sigma_{\theta j}, u_{ij}(x)$ ($j = 1, \dots, n$) be arbitrary generalized admissible variations on $x_1 x_2$. Then for arbitrary values of c and d ,*

$$(8.6) \quad |cu_{ij}(x_3^-) + du_{ij}(x_3^+)| = (c + d)^{n-1} [c |u_{ij}(x_3^-)| + d |u_{ij}(x_3^+)|] \\ (x_1 \leq x \leq x_2).$$

If x_3 does not define a corner of E_{12} , then equation (8.6) is trivial. If $x_3 = t_\theta$ ($\theta = 1, \dots, r$) the result may be proved in an elementary manner in view of the equations $X_i[\theta | \sigma, u] = 0$.

The following theorem is fundamental.

THEOREM 8.1. *Let E_{12} be a normal non-singular extremaloid which has $\Omega_0 \neq 0$ at its corners, and which is normal on every sub-interval $x_1 x_3$ ($x_1 < x_3 \leq x_2$). If $\tau_{\theta j}, \eta_{ij}(x), \xi_{ij}(x)$ ($j = 1, \dots, n$) are secondary extremaloids determined by the initial conditions*

$$(8.7) \quad \begin{aligned} \eta_{ih}(x_1) = 0, \quad \eta_{in}(x_1) = z_i(x_1) \quad (h = 1, \dots, n-1), \\ |\xi_{ih}(x_1)z_i(x_1)| \neq 0, \quad \xi_{in}(x_1) = 0, \end{aligned}$$

then a value $x_3 \neq x_1$ defines a point 3 conjugate to 1 on E_{12} if and only if

$$|\eta_{ij}(x_3^-)| \cdot |\eta_{ij}(x_3^+)| \leq 0.$$

On the assumption that E_{12} is normal on every sub-interval $x_1 x_3$ ($x_1 < x_3 \leq x_2$), it is seen that an arc $\eta_i = \eta_{ij}(x)b_j$ is identically zero on $x_1 x_3$ if and only if $b_j = 0$ ($j = 1, \dots, n$). It is also to be remarked that if for such a point x_3 there exist constants c, d, b_j such that

$$(8.8) \quad [c\eta_{ij}(x_3^-) + d\eta_{ij}(x_3^+)]b_j = 0 \quad (c + d \neq 0),$$

then $b_n = 0$. This result is immediate, since by Lemma 2.2 we have

$$0 = z_i(x_3) [c\eta_{ij}(x_3^-) + d\eta_{ij}(x_3^+)]b_j = (c + d)z_i(x_1)\eta_{ij}(x_1)b_j \\ = (c + d)z_i(x_1)z_i(x_1)b_n.$$

Now if x_3 is a point such that $|\eta_{ij}(x_3^-)| \cdot |\eta_{ij}(x_3^+)| \leq 0$, there exist constants c, d satisfying the relation

$$(8.9) \quad c |\eta_{ij}(x_3^-)| + d |\eta_{ij}(x_3^+)| = 0 \quad (cd \geq 0, c + d \neq 0).$$

As a consequence of Lemma 8.2 there then exist constants b_j not all zero, but with b_n necessarily zero, which satisfy equations (8.8) with the values c, d determined by (8.9). We have therefore established that if the de-

terminant $|\eta_{ij}(x)|$ vanishes or changes sign at $x = x_3$, then x_3 defines a point 3 conjugate to 1 on E_{12} .

Conversely, it is seen that every secondary extremaloid $\tau_\theta, \eta_i, \xi_i$ which satisfies the conditions $\eta_i(x_1) = 0$ is of the form

$$\eta_i = \eta_{ih}(x)b_h, \quad \xi_i = \xi_{ih}(x)b_h + z_i(x)b_n, \quad \tau_\theta = \tau_{\theta n}b_n.$$

If x_3 defines a point 3 conjugate to 1 on E_{12} there exist constants c, d such that $|c\eta_{ij}(x_3^-) + d\eta_{ij}(x_3^+)| = 0$, $cd \geq 0$, $c + d \neq 0$, and as a consequence of Lemma 8.2 we have $|\eta_{ij}(x_3^-)| \cdot |\eta_{ij}(x_3^+)| \leq 0$.

9. An auxiliary theorem. In this section we shall prove the following auxiliary theorem which is fundamental in the construction of a family of extremaloids for the problem of Mayer.

THEOREM 9.1. *Let E_{12} be a normal extremaloid, normal on every sub-interval x_1x_3 , which satisfies conditions III', IV', and for which $\Omega_0 < 0$ at the corner values $x = t_\theta$ ($\theta = 1, \dots, r$). Then there exists a set of n secondary extremaloids $S_{\theta j}, U_{ij}(x), V_{ij}(x)$ which are mutually conjugate in pairs and such that the determinant $|U_{ij}(x)|$ does not vanish or change sign on the interval $x_1 \leq x \leq x_2$.*

Preliminary to the proof of the above theorem we shall establish two lemmas. Consider a system of secondary extremaloids $\tau_{\theta j}, \eta_{ij}(x), \xi_{ij}(x)$ defined by initial conditions (8.7), and such that $\xi_{ih}(x_1)z_i(x_1) = 0$ ($h = 1, \dots, n-1$). The members of this family are then seen to be mutually conjugate in pairs. Furthermore, since E_{12} satisfies the condition IV', it follows that the determinant $|\eta_{ij}(x)|$ does neither vanish nor change sign on $x_1 < x \leq x_2$. If the matrices $\|\tau_{\theta j}\|, \|\eta_{ij}(x)\|, \|\xi_{ij}(x)\|$ are multiplied on the right by the inverse of the matrix $\|\eta_{ij}(x_2)\|$ a new conjugate system $\sigma_{\theta j}, u_{ij}(x), v_{ij}(x)$ is obtained which satisfies the conditions $u_{ij}(x_2) = \delta_{ij}, v_{ij}(x_2) = B_{ij} = v_{ji}(x_2)$ ($i, j = 1, \dots, n$). This conjugate system has the following properties:

LEMMA 9.1. *Let $\xi_s, \tau_\theta, \eta_i(x)$ be an arbitrary set of generalized admissible variations, and define on $x_1 < x \leq x_2$ functions $a_j(x)$ by the relation $\eta_i(x) = u_{ij}(x)a_j(x)$, where the functions $u_{ij}(x)$ belong to the conjugate set of secondary extremaloids $\sigma_{\theta j}, u_{ij}(x), v_{ij}(x)$ defined above. Then*

$$(9.1) \quad [\tau_\theta - \sigma_{\theta j}a_j(t_\theta^-)][\tau_\theta - \sigma_{\theta j}a_j(t_\theta^+)] \geq 0 \quad (\theta = 1, \dots, r),$$

$$(9.2) \quad 2H[\theta | \tau, \eta] - [\eta_i(x)v_{ij}(x)a_j(x)]_{t_\theta^-}^{t_\theta^+} + [\tau_\theta - \sigma_{\theta j}a_j(t_\theta^-)][\tau_\theta - \sigma_{\theta j}a_j(t_\theta^+)]\Omega_0(t_\theta) = 0.$$

The equations $X_i[\theta | \tau, \eta] = 0$, $X_i[\theta | \sigma_j, u_j] = 0$ are seen to imply that if for a given value of θ the functions $a_j(t_\theta^-) - a_j(t_\theta^+)$ are all zero then the quantity $\tau_\theta - \sigma_{\theta j} a_j(t_\theta^-)$ is also zero. It also follows that the functions $u_i(x | \theta) = u_{ij}(x) [a_j(t_\theta^-) - a_j(t_\theta^+)]$ satisfy the relations

$$[\tau_\theta - \sigma_{\theta j} a_j(t_\theta^-)] u_i(t_\theta^- | \theta) - [\tau_\theta - \sigma_{\theta j} a_j(t_\theta^+)] u_j(t_\theta^+ | \theta) = 0.$$

Relation (9.1) is then seen to follow as a consequence of the condition that E_{12} satisfies IV' .

To prove (9.2) it is only necessary to verify that the expression on the left-hand side of the equation is identical with the rather complicated quantity

$$\begin{aligned} & \{ \tau_\theta [F_{y_i}(t_\theta^-) + F_{y_i}(t_\theta^+)] - F_{y_i}(t_\theta^+) \sigma_{\theta k} a_k(t_\theta^+) - F_{y_i}(t_\theta^-) \sigma_{\theta k} a_k(t_\theta^-) \} X_i[\theta | \tau, \eta] \\ & + \{ a_j(t_\theta^-) + a_j(t_\theta^+) \} \{ -X_i[\theta | \tau, \eta] d_{\theta | i, j} - X_{2n+i}[\theta | \sigma_j, u_j, d_j] \eta_i(t_\theta^+) \\ & + X_{3n+1}[\theta | \sigma_j, u_j, d_j] \tau_\theta \} + a_j(t_\theta^-) \{ X_{n+i}[\theta | \sigma_j, u_j, d_j] \eta_i(t_\theta^-) \\ & - F_{y_i}(t_\theta^+) X_i[\theta | \sigma_j, u_j] \tau_\theta \} - \tau_\theta F_{y_i}(t_\theta^-) X_i[\theta | \sigma_j, u_j] a_j(t_\theta^+) \\ & + a_k(t_\theta^+) \{ [d_{\theta | i, k} + \sigma_{\theta k} F_{y_i}(t_\theta^+)] X_i[\theta | \sigma_j, u_j] \\ & + F_{y_i}(t_\theta^-) X_i[\theta | \sigma_k, u_k] \sigma_{\theta j} \\ & + u_{ij}(t_\theta^+) X_{2n+i}[\theta | \sigma_k, u_k, d_k] \\ & - X_{3n+1}[\theta | \sigma_k, u_k, d_k] \sigma_{\theta j} \} a_j(t_\theta^-). \end{aligned}$$

LEMMA 9.2. Suppose σ_θ , $u_i(x)$, $v_i(x)$ is a secondary extremaloid and there is a point x_3 on $x_1 \leq x_3 < x_2$ such that

$$(9.3) \quad cu_i(x_3^-) + du_i(x_3^+) = 0 \quad cd \geq 0, \quad c + d \neq 0.$$

Now define an arc η_i by the equations

$$\eta_i = 0 \quad (x_1 \leq x < x_3), \quad \eta_i = u_i(x) \quad (x_3 < x \leq x_2)$$

and determine associated constants ξ_s , τ_θ as follows: ξ_s ($s = 1, 2$) arbitrary; $\tau_\theta = 0$ if $t_\theta < x_3$, $\tau_\theta = \sigma_\theta$ if $t_\theta > x_3$; if x_3 coincides with a corner point t_{θ^*} of E_{12} then $\tau_{\theta^*} = c\sigma_{\theta^*}/(c + d)$. Then the set of generalized admissible variations ξ_s , τ_θ , $\eta_i(x)$ so defined satisfies the inequality

$$(9.4) \quad g_2[\xi, \tau, \eta] - \eta_i(x_2) B_{ij} \eta_j(x_2) - 2Q[\xi, \eta] \geq 0.$$

Consider first the case when $x_1 < x_3 < x_2$. It then follows by the Clebsch transformation of the second variation [III, p. 738] that

$$\begin{aligned} g_2[\xi, \tau, \eta] &= \eta_i(x_2) B_{ij} \eta_j(x_2) + \sum_{\theta=1}^r 2H[\theta | \tau, \eta] - \eta_i(x) v_{ij}(x) a_j(x) \Big|_{t_\theta^-}^{t_\theta^+} \\ &\quad + 2Q[\xi, \eta] + \int_{x_1}^{x_2} F_{y', v'}(\eta' - u'_{ik} a_k)(\eta'_j - u'_{j1} a_1) dx. \end{aligned}$$

The expression (9.4) is then seen to be non-negative in view of (9.1), (9.2), condition III', and the hypothesis that $\Omega_0 < 0$ at the corners of E_{12} . Suppose

now that $x_3 = x_1$. It is to be remarked first that we may assume without loss of generality that $\xi_i(x_1)z_i(x_1) = 0$, since this additional property may be obtained by adding a suitable multiple of the set $\xi_s = 0$, $\tau_\theta = 0$, $\eta_i \equiv 0$, $\xi_i(x) = z_i(x)$, and such modification does not change the value of the expression in (9.4). The modified secondary extremaloid, however, is seen to be expressible linearly in terms of the members of the family $\sigma_{\theta j}$, u_{ij} , v_{ij} , and by direct integration it is found that $g_2[\xi, \tau, \eta] - 2Q[\xi, \eta] = \eta_i(x_2)B_{ij}\eta_j(x_2)$. Hence the lemma is established.

In order to establish Theorem 9.1 it will now be proved that the set of mutually conjugate secondary extremaloids $S_{\theta j}$, $U_{ij}(x)$, $V_{ij}(x)$ having the initial values

$$(9.5) \quad U_{ij}(x_2) = \delta_{ij}, \quad V_{ij}(x_2) = D_{ij} = B_{ij} - \delta_{ij}$$

has the property that the determinant $|U_{ij}(x)|$ neither vanishes nor changes sign on $x_1 \leq x \leq x_2$. In the first place, $|U_{ij}(x_2)| = 1$. If now $|U_{ij}(x)|$ were to vanish or change sign for a value x_3 ($x_1 \leq x_3 < x_2$), there would exist constants a_j not all zero such that the secondary extremaloid $\sigma_\theta = S_{\theta j}a_j$, $u_i = U_{ij}a_j$, $v_i = V_{ij}a_j$ satisfied at x_3 the condition (9.3). Then consider the set ξ_s , τ_θ , η_i , ξ_i of Lemma 9.2. In case x_3 did not correspond to a corner of E_{12} , it could be proved by direct integration that

$$g_2[\xi, \tau, \eta] - \eta_i(x_2)B_{ij}\eta_j(x_2) - 2Q[\xi, \eta] = a_i D_{ij}a_j - a_i B_{ij}a_j = -a_i a_i < 0.$$

In case x_3 coincided with a corner value $x = t_{\theta^*}$ of E_{12} , then by direct integration and the use of relation (8.4) it would follow that

$$g_2[\xi, \tau, \eta] - \eta_i(x_2)B_{ij}\eta_j(x_2) - 2Q[\xi, \eta] \\ = a_i D_{ij}a_j - a_i B_{ij}a_j + cd\tau_{\theta^*}^2 \Omega_0(t_{\theta^*}) / (c + d)^2, \leq -a_i a_i < 0,$$

in view of the assumption that $\Omega_0 < 0$ at the corners of E_{12} . In either case we would have a contradiction to the result of Lemma 9.2. Hence the determinant $|U_{ij}(x)|$ neither vanishes nor changes sign on $x_1 \leq x \leq x_2$ and Theorem 9.1 is established.

10. Definition of a field and a fundamental sufficiency theorem.

A field with r discontinuities is defined to be a region \mathfrak{F} in xy -space containing only interior points, having associated with it r corner manifolds

$$(10.1) \quad \mathfrak{M}_\theta : x = T_\theta(a_1, \dots, a_n), \quad y_i = Y_{\theta|i}[T_\theta(a_1, \dots, a_n), a_1, \dots, a_n], \\ (\theta = 1, \dots, r)$$

and functions $p_i(x, y)$, $l_\alpha(x, y)$ ($i = 1, \dots, n$; $\alpha = 1, \dots, m$) with the following properties:

(a) the corner manifolds \mathfrak{M}_θ are non-singular, have no multiple points, do not intersect each other and each \mathfrak{M}_θ divides \mathfrak{F} into two parts;

(b) the functions $p_i(x, y)$, $l_a(x, y)$ are continuous and have continuous first derivatives between corner manifolds \mathfrak{M}_θ in \mathfrak{F} and approach finite limits on each side of each \mathfrak{M}_θ ;

(c) the two limits p_i^- and p_i^+ of the functions p_i at a point of a corner manifold \mathfrak{M}_θ are such that the sets p_i^- and p_i^+ always determine directions on the same side of \mathfrak{M}_θ and never tangent to it;

(d) the sets $[x, y, p(x, y)]$ for (x, y) in \mathfrak{F} are all admissible;

(e) the functions

$$F_{y'_i}[x, y, p(x, y), l(x, y)], \\ F[x, y, p(x, y), l(x, y)] - p_i(x, y)F_{y'_i}[x, y, p(x, y), l(x, y)]$$

are continuous in \mathfrak{F} ;

(f) the Hilbert integral

$$I^* = \int \{F[x, y, p, l]dx + (dy_i - p_i dx)F_{y'_i}[x, y, p, l]\}$$

formed with these functions is independent of the path in \mathfrak{F} . At a point (x, y) of a corner manifold of \mathfrak{F} it is to be understood that the set of values (x, y, p, l) appearing in the integrand of I^* are replaced by either the set (x, y, p^-, l^-) or the set (x, y, p^+, l^+) . From (e) it follows that the value of the integrand is the same for either set of limiting values.

It may be proved in the usual manner that in the field between corner manifolds \mathfrak{M}_θ the solutions of the differential equations

$$dy_i/dx = p_i(x, y)$$

are extremal arcs. Condition (e) above then implies that these may be pieced together to form extremaloids, which are called the extremaloids of the field. The value of I^* is seen to be zero along every extremaloid of the field.

Sufficient conditions for the problem of Mayer here discussed are obtained by the same method that Hestenes [IX] has used for the case of extremal arcs, and makes use of the following auxiliary Mayer problem. On the assumption that E_{12} is a normal extremaloid which satisfies the necessary condition I it is seen that the matrix

$$(10.2) \quad \begin{vmatrix} g_{x_1} & g_{y_{11}} & g_{x_2} & g_{y_{12}} \\ \psi_{\mu x_1} & \psi_{\mu y_{11}} & \psi_{\mu x_2} & \psi_{\mu y_{12}} \end{vmatrix}$$

has rank $p+1$. Let $\psi_v[x_1, y_1, x_2, y_2]$ ($v=p+1, \dots, 2n+1$) be $2n+1-p$ functions possessing continuous first and second partial derivatives in a neighborhood of the end-values $[x_{10}, y_{10}, x_{20}, y_{20}]$ belonging to E_{12} , vanishing at these values, and for these values having the determinant

$$(10.3) \quad \begin{vmatrix} g_{x_1} & g_{y_{11}} & g_{x_2} & g_{y_{12}} \\ \psi_{\rho x_1} & \psi_{\rho y_{11}} & \psi_{\rho x_2} & \psi_{\rho y_{12}} \end{vmatrix}$$

different from zero. The new set of end conditions $\psi_\rho = 0$ ($\rho = 1, \dots, 2n+1$) defines an auxiliary problem of the type discussed by Bliss and Hestenes [VIII], and a minimizing extremaloid E_{12} for the original Mayer problem is also a minimizing extremaloid for this auxiliary problem. Furthermore, if E_{12} is not only normal, but also normal on x_1x_2 , then it is normal with respect to the conditions $\psi_\rho = 0$ ($\rho = 1, \dots, 2n+1$) as defined above [see IX, p. 484].

For this auxiliary problem of Mayer we may prove by the same methods that Bliss and Hestenes have used for extremal arcs [VIII, § 7] the following theorem.

THEOREM 10.1. A FUNDAMENTAL SUFFICIENCY THEOREM. *Consider a normal extremaloid E_{12} which is an extremaloid of a field \mathfrak{F} . Suppose that the ends of E_{12} satisfy the conditions $\psi_\rho = 0$ and that there is a neighborhood N of these ends in $(x_1y_1x_2y_2)$ -space such that no other extremaloid of the field has ends in N satisfying the equations $\psi_\rho = 0$. If at each point of \mathfrak{F} the condition*

$$(10.5) \quad \mathcal{E}[x, y, p^-(x, y), l^-(x, y), y'] > 0$$

holds for every admissible set (x, y, y') such that $(x, y, y') \neq (x, y, p^-)$, $(x, y, y') \neq (x, y, p^+)$, then the neighborhood N can be so restricted that the inequality $g(C_{34}) > g(E_{12})$ is true for every admissible arc C_{34} in \mathfrak{F} with ends in N satisfying the conditions $\psi_\rho = 0$ and not identically with E_{12} .

Between corner manifolds the inequality (10.5) gives the usual strengthened form of the Weierstrass condition. On the corner manifolds this inequality implies in view of condition (e) above that we have also $\mathcal{E}[x, y, p^+(x, y), l^+(x, y), y'] > 0$ for every admissible set (x, y, y') such that $(x, y, y') \neq (x, y, p^-)$, $(x, y, y') \neq (x, y, p^+)$.

11. Sufficient conditions for a strong relative minimum. The construction of a field of extremaloids for the auxiliary Mayer problem defined in § 10 is embodied in the following theorem.

THEOREM 11.1. *Let E_{12} be a normal extremaloid that is normal on every sub-interval x_1x_3 ($x_1 < x_3 \leq x_2$), which satisfies conditions III', IV', and for which $\Omega_0 \neq 0$ at its corners. Then E_{12} is a member of an n -parameter family of extremaloids*

$$(11.1) \quad y_i = y_i(x, b_1, \dots, b_n), \quad z_i = z_i(x, b_1, \dots, b_n)$$

whose determinant $|y_{ib}|$ neither vanishes nor changes sign along E_{12} ; furthermore, E_{12} is an extremaloid of a field \mathfrak{F} simply covered by the family.

The explicit form of such a family may be given as follows. Let $B(b_1, \dots, b_n)$ be the function

$$B(b) = z_{i2}b_i + (1/2)D_{ij}(b_i - y_{i2})(b_j - y_{j2}),$$

where the constants D_{ij} are defined by equations (9.5) in terms of the end-values of the conjugate system of secondary extremaloids $S_{\theta ij}$, U_{ij} , V_{ij} determined in § 9 to satisfy the conditions of Theorem 9.1, and y_{i2} , z_{i2} are the end values at $x = x_{20}$ of the set $y_i(x)$, $z_i(x)$ defining E_{12} . When in equations (3.7) the set (x_0, y_0, z_0) is replaced by the set (x_{20}, b_i, B_{b_i}) an n -parameter family of extremaloids

$$(11.2) \quad y_i = y_i(x, x_{20}, b, B_b) = y_i(x, b), \quad z_i = z_i(x, x_{20}, b, B_b) = z_i(x, b)$$

is defined and contains E_{12} for $b_i = y_{i2}$. The multipliers $\lambda_a(x, b)$ and corner manifolds \mathfrak{M}_θ associated with this family are determined by equations (3.2) and (3.8). Furthermore, since each extremaloid (11.2) defined by parameter values b_i has on it the element (x_{20}, b_i, B_{b_i}) , it is seen that $y_{ib_j} = \delta_{ij}$, $z_{ib_j} = D_{ij}$ at $x = x_{20}$. Hence from Theorem 9.1 it follows that the determinant $|y_{ib_j}|$ neither vanishes nor changes sign along the extremaloid E_{12} of the family (11.2). This family, therefore, simply covers a neighborhood of E_{12} . In particular, let \mathfrak{F} be a neighborhood of E_{12} which is simply covered by the family, for which the portions of the corner manifolds \mathfrak{M}_θ of the family which lie in \mathfrak{F} satisfy condition (a) above, and such that if the parameter values of the extremaloid through a point (x, y) of \mathfrak{F} are denoted by $b_i(x, y)$, then the functions

$$(11.3) \quad p_i(x, y) = y_{ix}[x, b(x, y)], \quad l_a(x, y) = \lambda_a[x, b(x, y)]$$

have the sets $[x, y, p(x, y)]$ all admissible. The neighborhood \mathfrak{F} and the functions $p_i(x, y)$, $l_a(x, y)$ defined by (11.3) are seen to satisfy conditions (a)-(e) of § 10. Moreover, on the hyperplane $x = x_{20}$ the Hilbert integral I^* is expressible as

$$I^* = \int F_{y'} dy_i = \int B_b db_i = \int dB$$

and hence is independent of the path. The integral I^* may then be proved to be independent of the path in \mathfrak{F} by an argument that is a slight extension of that used to prove the corresponding result for the case of extremal arcs [see VIII, p. 323; also III, p. 733]. Hence \mathfrak{F} is a field as defined in § 10.

The proof of the following result is the same as for the case of extremal arcs [VIII, p. 323]:

THEOREM 11.2. *Let E_{12} be a normal extremaloid that is normal on x_1x_2 , and which is a member of an n -parameter family of extremaloids (11.1) whose determinant $|y_{i0}|$ neither vanishes nor changes sign along E_{12} . If the ends of E_{12} satisfy the conditions $\psi_\rho = 0$ ($\rho = 1, \dots, 2n+1$) of the auxiliary problem of § 10, then there is a neighborhood N of these ends in $(x_1y_1x_2y_2)$ -space such that E_{12} is the only extremaloid of the family with ends in N satisfying the conditions $\psi_\rho = 0$.*

The following theorem, which states sufficient conditions for a proper strong relative minimum in the auxiliary Mayer problem formulated in § 10, is seen to be an immediate consequence of the preceding theorems.

THEOREM 11.3. *Let E_{12} be an arc which satisfies the following hypotheses:*

H_1). E_{12} is an admissible arc with r corners and with ends satisfying conditions $\psi_\rho = 0$ ($\rho = 1, \dots, 2n+1$) of the auxiliary problem of § 10.

H_2). E_{12} is normal, normal on every sub-interval x_1x_3 of x_1x_2 , and satisfies conditions (I), (III'), (IV').

H_3). $\Omega_0 \neq 0$ at the corners on E_{12} .

H_4). The extremaloids of the family (11.1) determined by Theorem 11.1 satisfy condition II'_0 for values of the parameters b_i in a neighborhood of the set $b_i = b_{i0}$ defining E_{12} .

Then there are neighborhoods \mathfrak{F} of E_{12} in xy -space and N of the ends of E_{12} in $(x_1y_1x_2y_2)$ -space such that the inequality $g(C_{34}) > g(E_{12})$ holds for every admissible arc C_{34} in \mathfrak{F} with ends in N satisfying the conditions $\psi_\rho = 0$ and not identical with E_{12} .

The proof of sufficient conditions for a strong relative minimum in the general problem of Mayer follows the methods introduced by Hestenes for the case of extremal arcs. Use is made of a family of fields depending upon $n-1$ parameters which includes for special values of the parameters the field introduced above. In the following discussion it will be assumed that E_{12} satisfies the hypotheses of Theorem 11.1. Let $B_h(b_1, \dots, b_n)$ ($h = 1, \dots, n-1$) be functions of class C'' and such that the determinant $|B_b, B_{hb}|$ is different from zero for $b_i = b_{i0}$. Then the $(2n-1)$ -parameter family of extremaloids

$$(11.4) \quad \begin{aligned} y_i &= y_i(x, x_{20}, b, B_b + c_h B_{hb}) = y_i(x, b, c), \\ z_i &= z_i(x, x_{20}, b, B_b + c_h B_{hb}) = z_i(x, b, c) \end{aligned} \quad (x_1 \leq x \leq x_2)$$

contains E_{12} for $(x_1, x_2, b, c) = (x_{10}, x_{20}, b_0, 0)$. For this $(2n-1)$ -parameter family the determinant

$$\begin{vmatrix} y_{ib_j} & y_{ic_h} & 0 \\ z_{ib_j} & z_{ic_h} & z_i \end{vmatrix}$$

is found to be different from zero for the values $(x, b, c) = (x_{20}, b_0, 0)$.

Since the determinant $|y_{ib_j}|$ belonging to the family (11.2) neither vanishes nor changes sign along E_{12} , the determinant $|y_{ib_j}(x, b, c)|$ belonging to the family (11.4) has the same property. Hence the system of equations

$$(11.5) \quad y_i = y_i(x, b, c)$$

has a unique solution

$$b_i = b_i(x, y, c)$$

in a neighborhood \mathfrak{D} of the values (x, y, c) belonging to E_{12} . We shall suppose that \mathfrak{D} is so restricted that for parameter values b_i, c_h sufficiently near to the set b_{i0}, c_{h0} defining E_{12} the portions of the corner manifolds \mathfrak{M}_θ for the family (11.4) which lie in \mathfrak{D} satisfy condition (a) of § 10; moreover, if

$$(11.6) \quad \begin{aligned} p_i(x, y, c) &= y_{ix}[x, b(x, y, c), c], \\ l_a(x, y, c) &= l_a[x, b(x, y, c), c], \end{aligned}$$

then the sets $[x, y, p(x, y, c)]$ are all admissible.

By the same argument as used above we have that on the hyper-plane $x = x_{20}$ in xy -space the Hilbert integral I^* is independent of the path for fixed values of the parameters c_h . It follows that for each set c_h the region \mathfrak{F} of points (x, y) whose elements are all in \mathfrak{D} forms a field with slope functions and multipliers defined by equations (11.6). If in each of these fields the Weierstrass \mathcal{E} -function with arguments p_i, l_a defined by (11.6) is positive for every admissible set (x, y, y') distinct from (x, y, p^+) and (x, y, p^-) , then the methods used by Hestenes [IX] for the case of extremal arcs apply to give sufficient conditions for a proper strong relative minimum. We have the following theorem:

THEOREM 11.4. *Let E_{12} be an arc which satisfies the hypotheses (H_1) , (H_2) , (H_3) of Theorem 11.3 and the following strengthened form of hypothesis (H_4) of that theorem.*

H_4^). The extremaloids of the family (11.4) satisfy condition II'_0 for values of the parameters b_i, c_h in a neighborhood of the values $b_i = b_{i0}$, $c_h = c_{h0}$ defining E_{12} .*

Then there are neighborhoods \mathfrak{F} of E_{12} in xy -space and N of the ends of E_{12} in (x_1, y_1, x_2, y_2) -space such that the inequality $g(C_{34}) > g(E_{12})$ holds for every admissible arc C_{34} in \mathfrak{F} with ends in N satisfying the conditions $\phi_\mu = 0$ and not identical with E_{12} .

Suppose now that the functions ψ_μ are continuous at every pair of distinct or coincident points in a neighborhood of those belonging to E_{12} . If the ends of E_{12} are the only pair of distinct or coincident points on E_{12} satisfying the conditions $\psi_\mu = 0$, then for every $\delta > 0$ there exists a neighborhood \mathfrak{F} of E_{12} such that every pair of points in \mathfrak{F} satisfying the equations $\psi_\mu = 0$ necessarily has one of the points of the pair in each of the two δ -neighborhoods of the ends of E_{12} . Hence by suitably restricting the neighborhood \mathfrak{F} of E_{12} we obtain the following corollary.

COROLLARY. *Let E_{12} be an arc satisfying the hypotheses of Theorem 11.4. If further the ends of E_{12} are the only pair of distinct or coincident points on E_{12} satisfying the conditions $\psi_\mu = 0$, then there is a neighborhood \mathfrak{F} of E_{12} in xy -space such that the inequality $g(C_{34}) > g(E_{12})$ holds for every admissible arc C_{34} in \mathfrak{F} with ends satisfying the conditions $\psi_\mu = 0$ and not identical with E_{12} .*

In conclusion, it is to be remarked that in Theorems 11.3 and 11.4 the condition II'_0 has been assumed to be satisfied along the members of certain particular families of extremaloids. Let E denote an arbitrary extremaloid with equations (2.1) which has a corner for $x = t$, and for which $\Omega_0 < 0$ at this corner. The results of § 4 imply that there are sets of elements (x, y, y', Y', λ) for which the set $[x, y, y', \lambda]$ is in a given neighborhood of either the set $[t, y(t), y'(t-), \lambda(t-)]$ or the set

$$[t, y(t), y'(t+), \lambda(t+)],$$

and for which $\mathcal{E}[x, y, y', Y', \lambda] < 0$. Consequently, for the non-parametric problem it seems necessary to phrase the strengthened Weierstrass condition in the manner indicated above. In this respect, therefore, the statement of sufficient conditions for a strong relative minimum is more cumbersome for non-parametric problems than for the corresponding parametric problems [see IV, § 5].

Remarks. [Added to proof sheets, November 24, 1934.]

In a recent paper Hestenes has proved sufficiency theorems for an extremal arc in the problem of Bolza assuming no normality conditions aside from the existence of multipliers of the form $\lambda_0 = 1, \lambda_\alpha(x)$, by modifying the usual form of the Mayer condition. Hestenes' methods may be extended to establish corresponding theorems for the problem of discontinuous solutions here studied.* Indeed, such sufficiency theorems may be established by the same

* *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 793-818.

method as used above. This fact is a consequence of a result that has recently been established independently by Morse and the author.*

BIBLIOGRAPHY

- I. Bliss, "The problem of Mayer with variable end points," *Transactions of the American Mathematical Society*, vol. 19 (1918), pp. 305-314.
- II. Cope, "An analogue of Jacobi's condition for the problem of Mayer with variable end points," Dissertation, The University of Chicago (1927).
- III. Bliss, "The problem of Lagrange in the calculus of variations," *American Journal of Mathematics*, vol. 52 (1930), pp. 673-744.
- IV. Graves, "Discontinuous solutions in space problems of the calculus of variations," *American Journal of Mathematics*, vol. 52 (1930), pp. 1-28.
- V. Graves, "Discontinuous solutions in the calculus of variations," *Bulletin of the American Mathematical Society*, vol. 36 (1930), pp. 831-846.
- VI. Hefner, "The condition of Mayer for discontinuous solutions of the Lagrange problem," *Contributions to the Calculus of Variations 1931-1932*—Theses submitted to the Department of Mathematics of the University of Chicago, pp. 95-130.
- VII. Reid, "Boundary value problems associated with the calculus of variations," *American Journal of Mathematics*, vol. 54 (1932), pp. 769-790.
- VIII. Bliss and Hestenes, "Sufficient conditions for a problem of Mayer in the calculus of variations," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 305-326.
- IX. Hestenes, "Sufficient conditions for the general problem of Mayer with variable end points," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 479-490.
- X. Reid, "Analogues of the Jacobi condition for the problem of Mayer in the calculus of variations," *Annals of Mathematics*, vol. 35 (1934), pp. 836-848.

UNIVERSITY OF CHICAGO,
CHICAGO, ILLINOIS.

* For abstracts of results, see *Bulletin of the American Mathematical Society*, vol. 40 (1934), p. 665 and p. 666. The central theorem concerning secondary extremal arcs which is involved has also been established by Hestenes.

ON THE INVERSION FORMULA FOR FOURIER-STIELTJES TRANSFORMS IN MORE THAN ONE DIMENSION.

By E. K. HAVILAND.

An inversion formula for Fourier-Stieltjes transforms in one dimension, given by P. Lévy,* has been extended to more than one dimension by V. Romanovsky † in a formal treatment. It is the purpose of the present paper to give an exact proof of this formula. For simplicity, the proof is given for the case of two dimensions, but its extension to n -dimensions is given by precisely the same methods.

It is known ‡ that if a function $f(x, y)$ is

- (i) bounded throughout the entire (xy) -plane,
- (ii) of bounded variation in every finite region of the plane,
- (iii) absolutely integrable over the entire (xy) -plane,

then to the formula

$$(1) \quad E(s, t) = (2\pi)^{-2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \exp [i(sx + ty)] dx dy$$

there corresponds the inverse formula:

$$(2) \quad f(u, v) = \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \right) E(s, t) \exp [-i(us + vt)] ds dt,$$

where the latter double integral is to be considered as a Cauchy principal value, as indicated by the parenthesis, and (u, v) is any continuity point of $f(u, v)$.

Accordingly, we proceed to show that if $\phi(E)$ is a distribution function,§ and $F(x, y)$ is the corresponding point function,¶ then ||

* P. Lévy, *op. cit.*, pp. 166-167. References are collected at the end of the paper.

† V. Romanovsky, *loc. cit.*, pp. 36-40.

‡ Cf., e. g., V. Romanovsky, *loc. cit.*, pp. 36-38; also S. Bochner, *op. cit.*, p. 202, Theorem 66. These authors do not give the theorem precisely in the above form, but a proof of the latter may be obtained by the use of their methods together with certain analogues of the Second Theorem of the Mean given by W. H. Young, *loc. cit.*, p. 290, equations (23), (24), (25) and (26).

§ The monotone absolutely additive set function $\phi(E)$ is said to be a distribution function if $0 \leq \phi(E) \leq 1$ and $\phi(S) = 1$, where S denotes the whole (xy) -plane. Cf. E. K. Haviland, *loc. cit.*, p. 627.

¶ Cf. J. Radon, *loc. cit.*, pp. 1304-1305.

|| $R + P_{\xi\eta}$ represents the vector addition of the quadrant R and the point $P_{\xi\eta} = (\xi, \eta)$.

$$\begin{aligned}
 (3) \quad g(x, y) &= G(x, y; \xi, \eta) \\
 &= \phi(R + P_{\xi\eta}) - \phi(R + P_{\xi 0}) - \phi(R + P_{0\eta}) + \phi(R) \\
 &= F(x + \xi, y + \eta) - F(x + \xi, y) - F(x, y + \eta) + F(x, y),
 \end{aligned}$$

where $R: (-\infty < u < x; -\infty < v < y)$, as a function of x and y , fulfills conditions (i), (ii), (iii) above.

That $G(x, y; \xi, \eta)$ is bounded for all (x, y) , ξ and η being fixed, is seen at once from the fact that ϕ is a distribution function, in which case $|G| < 4$.

That $G(x, y; \xi, \eta)$ is of bounded variation with respect to (x, y) may be seen by examining its four terms separately. In the first place, $F(x + \xi, y + \eta)$ is a monotone function of (x, y) ; for if $x_2 > x_1$, $y_2 > y_1$, and $R_{ij}: (-\infty < x < x_i; -\infty < y < y_j)$,

$$\begin{aligned}
 &F(x_2 + \xi, y_2 + \eta) - F(x_2 + \xi, y_1 + \eta) - F(x_1 + \xi, y_2 + \eta) + F(x_1 + \xi, y_1 + \eta) \\
 &= \phi(R_{22} + P_{\xi\eta}) - \phi(R_{21} + P_{\xi\eta}) - \phi(R_{12} + P_{\xi\eta}) + \phi(R_{11} + P_{\xi\eta}) \\
 &= \phi(E_1 + P_{\xi\eta}) \geq 0, \text{ where } E_1: (x_1 \leq x < x_2; y_1 \leq y < y_2),
 \end{aligned}$$

since ϕ is a distribution function. Similarly, $F(x + \xi, y)$, $F(x, y + \eta)$ and $F(x, y)$ are bounded monotone functions of (x, y) , so that $G(x, y; \xi, \eta)$ is of bounded total variation throughout the (x, y) -plane.*

To prove $G(x, y; \xi, \eta)$ absolutely integrable throughout the (x, y) -plane for fixed ξ and η , we observe that †

$$\begin{aligned}
 &\int_a^b \int_c^d |G(x, y; \xi, \eta)| \, dx dy = \pm \int_a^b \int_c^d G(x, y; \xi, \eta) \, dx dy \\
 &= \pm \int_a^b \int_c^d [F(x + \xi, y + \eta) - F(x + \xi, y) - F(x, y + \eta) + F(x, y)] \, dx dy \\
 &= \pm \left\{ \int_{a+\xi}^{b+\xi} \int_{c+\eta}^{d+\eta} F(x, y) \, dx dy - \int_{a+\xi}^{b+\xi} \int_c^d F(x, y) \, dx dy \right. \\
 &\quad \left. - \int_a^b \int_{c+\eta}^{d+\eta} F(x, y) \, dx dy + \int_a^b \int_c^d F(x, y) \, dx dy \right\} \\
 &= \pm \left\{ \int_{a+\xi}^{b+\xi} \left[\int_{c+\eta}^{d+\eta} - \int_c^d \right] F(x, y) \, dx dy - \int_a^b \left[\int_{c+\eta}^{d+\eta} - \int_c^d \right] F(x, y) \, dx dy \right\} \\
 &= \pm \left\{ \int_{a+\xi}^{b+\xi} \left[\int_a^{d+\eta} - \int_c^{c+\eta} \right] F(x, y) \, dx dy \right. \\
 &\quad \left. - \int_a^b \left[\int_a^{d+\eta} - \int_c^{c+\eta} \right] F(x, y) \, dx dy \right\}
 \end{aligned}$$

* A statement by B. H. Camp, *loc. cit.*, p. 45 that a bounded monotone function is not necessarily of bounded variation in the sense of Hardy or Radon is incorrect.

† That the double and iterated integrals given below exist and are equivalent may be seen by consulting, e. g., E. W. Hobson, *op. cit.*, pp. 481 and 483.

$$\begin{aligned}
&= \pm \left\{ \int_{a+\xi}^{b+\xi} \int_0^\eta [F(x, y+d) - F(x, y+c)] dx dy \right. \\
&\quad \left. - \int_a^b \int_0^\eta [F(x, y+d) - F(x, y+c)] dx dy \right\} \\
&= \pm \left\{ \left[\int_{a+\xi}^{b+\xi} - \int_a^b \right] \int_0^\eta [F(x, y+d) - F(x, y+c)] dx dy \right\} \\
&= \pm \left\{ \left[\int_b^{b+\xi} - \int_a^{a+\xi} \right] \int_0^\eta [F(x, y+d) - F(x, y+c)] dx dy \right\} \\
&= \pm \int_0^\xi \int_0^\eta [F(x+b, y+d) - F(x+b, y+c) \\
&\quad - F(x+a, y+d) + F(x+a, y+c)] dx dy.
\end{aligned}$$

But as $F(x, y)$ is the point function associated with the distribution function $\phi(E)$, the integrand $\rightarrow 0$ uniformly with respect to (x, y) in $(0 \leq x \leq \xi; 0 \leq y \leq \eta)$ as $a, b \rightarrow +\infty$ or $-\infty$, c and d being arbitrary; or as $c, d \rightarrow +\infty$ or $-\infty$, a and b being arbitrary. Hence

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |G(x, y; \xi, \eta)| dx dy < +\infty, \quad \text{q. e. d.}$$

The Fourier-Stieltjes transform $\Lambda(s, t; \phi)$ of the distribution function $\phi(E)$ is given by*

$$\begin{aligned}
\Lambda(s, t; \phi) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp[isx + ty] d_{xy} \phi(E) \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp[isx + ty] d_{xy} F(x, y),
\end{aligned}$$

where $F(x, y)$ is the point function corresponding to the absolutely additive set function ϕ and s and t are real. It follows that if $s \neq 0$, $t \neq 0$,

$$\begin{aligned}
&\Lambda(s, t; \phi) (is)^{-1} (it)^{-1} [\exp(-is\xi) - 1] [\exp(-it\eta) - 1] \\
&= - (st)^{-1} \left\{ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp[i\{s(x-\xi) + t(y-\eta)\}] d_{xy} \phi(E) \right. \\
&\quad - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp[i\{s(x-\xi) + ty\}] d_{xy} \phi(E) \\
&\quad \left. - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp[isx + t(y-\eta)] d_{xy} \phi(E) \right.
\end{aligned}$$

* The integral is the limit as the rectangle $R \rightarrow S$, the entire (x, y) -plane, of $\iint_R \exp[isx + ty] d_{xy} \phi(E)$, the latter being defined by J. Radon, *loc. cit.*, pp. 1322-1324. The second and third members of this equation represent merely different notations for the same integral.

$$\begin{aligned}
& + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp [i(sx + ty)] d_{xy} \phi(E) \} \\
= & - (st)^{-1} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp [i(sx + ty)] d_{xy} [\phi(E + P_{\xi\eta}) \\
& - \phi(E + P_{\xi 0}) - \phi(E + P_{0\eta}) + \phi(E)],
\end{aligned}$$

so

$$\begin{aligned}
(4) \quad \Lambda(s, t; \phi) (is)^{-1} (it)^{-1} [\exp(-is\xi) - 1] [\exp(-it\eta) - 1] \\
= - (st)^{-1} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp [i(sx + ty)] d_{xy} g(x, y)
\end{aligned}$$

by the preceding footnote and by equation (3). If we set

$$(5) \quad \exp [i(sx + ty)] = f(x, y),$$

we obtain by the partial integration formula of W. H. Young: *

$$\begin{aligned}
(6) \quad (st)^{-1} \int_a^b \int_c^d g(x, y) d_{xy} f(x, y) & - (st)^{-1} \int_a^b \int_c^d f(x, y) d_{xy} g(x, y) \\
= (st)^{-1} \{ & \int_a^b f(x, c) d_x g(x, c) + \int_c^d f(a, y) d_y g(a, y) \\
& - \int_a^b f(x, d) d_x g(x, d) - \int_c^d f(b, y) d_y g(b, y) \\
& + [f(b, d)g(b, d) - f(b, c)g(b, c) - f(a, d)g(a, d) + f(a, c)g(a, c)] \}.
\end{aligned}$$

We shall show that as $a, c \rightarrow -\infty$ and $b, d \rightarrow +\infty$, all terms on right of the preceding equation vanish. By the one-dimensional partial integration formula:

$$\begin{aligned}
(st)^{-1} \int_a^b f(x, c) d_x g(x, c) & = - (st)^{-1} \int_a^b g(x, c) d_x f(x, c) \\
& + (st)^{-1} [g(x, c) f(x, c)]_{x=a}^{x=b}.
\end{aligned}$$

Now from (3) it is seen that

$$(7) \quad g(x, y) = G(x, y; \xi, \eta) = \pm \phi(Q + P_{xy}),$$

where $Q: (0 \leq u < \xi; 0 \leq v < \eta)$ or $(\xi \leq u < 0; \eta \leq v < 0)$ or $(0 \leq u < \xi; \eta \leq v < 0)$ or $(\xi \leq u < 0; 0 \leq v < \eta)$, depending on the signs of ξ and η . In the two former cases, $g(x, y)$ is non-negative; in the two latter, it is non-positive. Since ϕ is a distribution function, there exists a rectangle J so large that $0 \leq \phi(S - J) < \epsilon/4$, where S is the entire (x, y) -plane. Accordingly, if (x, y) is so distant that $Q + P_{xy}$ lies outside J ,

* Cf. W. H. Young, *loc. cit.*, p. 282.

$0 \leq \phi(Q + P_{xy}) < \epsilon/4$; from which fact it follows, as $|f(x, y)| = 1$, that there is an $M_1 > 0$ so large that if $|a|, |b|, |c|, |d| > M_1$,

$$\begin{aligned} & |st|^{-1} | [g(x, c)f(x, c)]_{x=a}^{x=b} | \\ &= |st|^{-1} | G(b, c; \xi, \eta)f(b, c) - G(a, c; \xi, \eta)f(a, c) | < \epsilon/2. \end{aligned}$$

Again,

$$\begin{aligned} & | (st)^{-1} \int_a^b g(x, c) dx f(x, c) | \leq |t|^{-1} \int_a^b |G(x, c; \xi, \eta)| dx \\ &= \pm t^{-1} \int_a^b [F(x + \xi, c + \eta) - F(x + \xi, c) - F(x, c + \eta) + F(x, c)] dx \\ &= \pm t^{-1} \left[\int_{a+\xi}^{b+\xi} - \int_a^b \right] [F(x, c + \eta) - F(x, c)] dx \\ &= \pm t^{-1} \left[\int_b^{b+\xi} - \int_a^{a+\xi} \right] [F(x, c + \eta) - F(x, c)] dx \\ &= \pm t^{-1} \int_0^\xi [F(b + x, c + \eta) - F(b + x, c) \\ &\quad - F(a + x, c + \eta) + F(a + x, c)] dx. \end{aligned}$$

Thus, as t and η are fixed, the integrand can, for all x in $[0, \xi]$ be made arbitrarily small by taking $|c|$ sufficiently large. Consequently, as $t \neq 0$,

$$| (st)^{-1} \int_a^b g(x, c) dx f(x, c) | < \epsilon/2 \text{ if } |c| > M_2, \text{ say.}$$

Therefore if $|a|, |b|, |c|, |d|$ are all $> \max(M_1, M_2)$,

$$| (st)^{-1} \int_a^b f(x, c) dx g(x, c) | < \epsilon.$$

Precisely similar reasoning shows that

$$| (st)^{-1} \int_a^b f(x, d) dx g(x, d) | < \epsilon, \quad | (st)^{-1} \int_c^d f(a, y) dy g(a, y) | < \epsilon$$

and $| (st)^{-1} \int_c^d f(b, y) dy g(b, y) | < \epsilon$ if $|a|, |b|, |c|, |d|$ are sufficiently large. Finally, for fixed $s, t \neq 0$,

$$\begin{aligned} & | (st)^{-1} \{ f(b, d)g(b, d) - f(b, c)g(b, c) - f(a, d)g(a, d) + f(a, c)g(a, c) \} | \\ & \leq |st|^{-1} \{ |g(b, d)| + |g(b, c)| + |g(a, d)| + |g(a, c)| \}. \end{aligned}$$

From equation (7) it can be seen that the foregoing expression may be made arbitrarily small by taking $|a|, |b|, |c|, |d|$ sufficiently large. Hence for all fixed $s, t \neq 0$, the right-hand side of (6) vanishes as $a, c \rightarrow -\infty$; $b, d \rightarrow +\infty$. The left-hand side then becomes

$$(st)^{-1} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) d_{xy} f(x, y) = (st)^{-1} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) d_{xy} g(x, y),$$

or by virtue of (4) and (5):

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) \exp[i(sx + ty)] dx dy \\ + (is)^{-1}(it)^{-1} \Lambda(s, t; \phi) [\exp(-is\xi) - 1] [\exp(-it\eta) - 1].$$

As the limit as $s \rightarrow 0$ and/or $t \rightarrow 0$ of the left-hand side of (6) exists, the same is true of the right-hand side, so that we have for any fixed s, t :

$$(2\pi)^{-2}(is)^{-1}(it)^{-1} \Lambda(s, t; \phi) [\exp(-is\xi) - 1] [\exp(-it\eta) - 1] \\ = (2\pi)^{-2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) \exp[i(sx + ty)] dx dy.$$

Then by (2):

$$(8) \quad g(u, v) = (2\pi)^{-2} \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \right) (is)^{-1}(it)^{-1} \Lambda(s, t; \phi) \\ [\exp(-is\xi) - 1] [\exp(-it\eta) - 1] \exp[-i(us + vt)] ds dt,$$

where $g(u, v) = G(u, v; \xi, \eta) = \phi(Q + P_{uv})$. This is the required inversion formula. It has been proved under the assumption that (u, v) is a continuity point of g , and the latter will certainly be the case if $Q + P_{uv}$ is a non-singular* rectangle of ϕ . Moreover, it is known that the non-singular rectangles of ϕ are everywhere dense, so that ϕ is essentially determined by the inversion formula.*

With the help of (8) and the multiplication rule for Fourier-Stieltjes transforms,† one can prove the Continuity Theorem for Fourier-Stieltjes transforms:‡ If $\{\Lambda_n(s, t) = \Lambda(s, t; \phi_n)\}$ is a sequence of Fourier-Stieltjes transforms of distribution functions such that the sequence converges uniformly in every finite rectangle of the st -plane to a function $\Lambda(s, t)$, then $\Lambda(s, t)$ is the Fourier-Stieltjes transform of a distribution function ϕ and $\phi_n \rightarrow \phi$ on all non-singular rectangles of the latter.

* Cf. E. K. Haviland, *loc. cit.*, p. 628.

† For a proof of this rule, cf. E. K. Haviland, *loc. cit.*, p. 651, Theorem V. In the proof of this theorem, it is to be noted that the last line on p. 652 should read

$$" \left| \sum_{k=1}^N \{\psi(R_k) - \sum_{j=1}^m \phi_1(R_k - P_j) \phi_2(E_j)\} \right| < \epsilon/2 "$$

and correspondingly the summation with respect to k at the top of p. 653 should be taken beneath the absolute value sign. It is to be noted that $\sum_{j=1}^m E_j$ need not extend beyond J .

‡ The one-dimensional case of this theorem is treated by P. Lévy, *op. cit.*, the two-dimensional case by V. Romanovsky, *loc. cit.*, p. 41 and by S. Bochner, *loc. cit.*, II, p. 403.

Finally, if $\Lambda(s, t; \phi)$ is absolutely integrable, it is possible to differentiate with respect to ξ and η beneath the integral sign in (8) and the resulting second mixed derivative is a continuous function of ξ and η in any finite region of the $(\xi\eta)$ -plane. As

$$\phi(Q + P_{uv}) = F(u + \xi, v + \eta) - F(u + \xi, v) - F(u, v + \eta) + F(u, v),$$

we may then infer, on setting $u = v = 0$, that $\phi(E)$ is absolutely continuous and that the point function $F(x, y)$ corresponding to it possesses a continuous mixed derivative $\partial^2 F / \partial x \partial y$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

S. Bochner, I.: *Vorlesungen ueber Fouriersche Integrale*. (Leipzig, 1932); II. "Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse." *Mathematische Annalen*, vol. 108 (1933), pp. 378-410.

B. F. Camp, "Singular multiple integrals with applications to series," *Transactions of the American Mathematical Society*, vol. 14 (1913), pp. 42-64.

E. K. Haviland, "On the theory of absolutely additive distribution functions," *American Journal of Mathematics*, vol. 56 (1934), pp. 625-658.

E. W. Hobson, *Theory of functions of a real variable*, vol. 1 (Cambridge, 1921).

P. Lévy, *Calcul des probabilités* (Paris, 1925).

J. Radon, "Theorie und Anwendungen der absolut additiven Mengenfunktionen," *Sitzungsberichte der mathematischen-naturwissenschaftlichen Klasse der Kaiserl. Akademie zu Wien*, vol. 122 (1913), pp. 1295-1438.

V. Romanovsky, "Sur un théorème limite du calcul des probabilités," *Recueil Mathématique de la Société Mathématique de Moscou*, vol. 36 (1929), pp. 36-64.

W. H. Young, "On multiple integration by parts and the second theorem of the mean," *Proceedings of the London Mathematical Society*, ser. 2, vol. 16 (1917), pp. 273-293.

A NOTE ON THE DISTRIBUTION OF THE ZEROS OF THE ZETA-FUNCTION.

By AUREL WINTNER.

The results of Littlewood* on the remainder term of the prime number theorem depend upon a close connection between the distribution of the prime numbers on the one hand and the *cyclical* distribution of the sequence

$$(1) \quad \gamma_1, \gamma_2, \dots \quad (0 < \gamma_1 \leq \gamma_2 \leq \dots)$$

on the other hand, where $\rho_1 = \beta_1 + i\gamma_1$, $\rho_2 = \beta_2 + i\gamma_2$, \dots denote the zeros of $\zeta(s)$ in the upper half-plane. While for the purposes of Littlewood only a consequence of the Dirichlet approximation theorem is relevant, it seems to be worth while to determine the asymptotic cyclical distribution of (1). The object of the present note is to point out the fact that *the sequence (1) is asymptotically equidistributed † to modulus c*, where c is any non-vanishing real number. In order to simplify the formulae we shall choose $c = 1$; it will be clear that the proof holds for any c .

We have to prove that

$$(2) \quad K(n; x)/n \rightarrow x \quad (n \rightarrow \infty),$$

where x is any point of the interval $0 \leq x < 1$ and $K(n; x)$ denotes the number of values k satisfying both inequalities

$$(3) \quad k \leq n, \quad \gamma_k \leq x.$$

The truth of (2) would follow in virtue of the Riemann-Mangoldt asymptotic formula from a known general lemma ‡ if we knew the sequence of the differences $\gamma_{k+1} - \gamma_k$ to be a monotone sequence. In reality, we do not even know that $\gamma_{k+1} - \gamma_k = 0$ is impossible, and the numerical results of Gram § show that the sequence of the differences $\gamma_{k+1} - \gamma_k$ is *not* monotone, at least

* Cf. A. E. Ingham, "The distribution of prime numbers," *Cambridge Tracts*, No. 30, London, 1932, pp. 96-104.

† "Gleichverteilt" in the sense of Weyl. Cf. G. Pólya and G. Szegő, *Aufgaben und Lehrsätze*, Berlin (1925), vol. I, pp. 70-73.

‡ *Ibid.*, p. 72.

§ J. P. Gram, "Note sur les zéros de la fonction $\zeta(s)$ de Riemann," *Acta Mathematica*, vol. 27 (1903), p. 297.

if small values of k are not discarded. For the proof of (2) one does not need, however, anything but the Riemann-Mangoldt asymptotic formula in its roughest form, viz.

$$(4) \quad N(T) \sim L(T/2\pi) \quad (T \rightarrow \infty),$$

where

$$(5) \quad L(T) = T \log T$$

and $N(T)$ denotes the number of values k satisfying the inequality

$$(6) \quad \gamma_k \leq T.$$

By the definitions (3) and (6) of $K(n; x)$ and $N(T)$ we have the identity

$$K(N(T); x) = N(1+x) - N(1) + N(2+x) - N(2) + \cdots \\ + N([T]-1+x) - N([T]-1) + N(\min\{T, [T]+x\}) - N([T])$$

so that

$$(7) \quad K(N(T); x) = \sum_{k=1}^{[T]} \{N(k+x) - N(k)\} + o(N(T))$$

in virtue of (4) and (5). Now (4) implies by a well known limit theorem that

$$\sum_{k=1}^{[T]} \{N(k+x) - N(k)\} \sim \sum_{k=1}^{[T]} \{L((k+x)/2\pi) - L(k/2\pi)\}$$

for every fixed x . This is easily seen to be

$$\sim \int_1^T \{L((x+k)/2\pi) - L(k/2\pi)\} dk,$$

hence $\sim xL(T/2\pi)$ in virtue of l'Hôpital's rule. Thus the sum $\sum_{k=1}^{[T]}$ occurring in (7) is $\sim xL(T/2\pi)$, hence $= xN(T) + o(N(T))$ in virtue of (4). Consequently $K(N(T); x) \sim xN(T)$, which is the same thing as (2).

THE JOHNS HOPKINS UNIVERSITY.

PARALLELISM AND EQUIDISTANCE IN RIEMANNIAN GEOMETRY.

By R. M. PETERS.

It is the purpose of this paper to generalize to Riemannian spaces the concepts and some of the results given by Graustein * for classical differential geometry. The angular spread, measure of the deviation from parallelism, in the sense of Levi-Civita, of one of two families of curves on a two-dimensional surface with respect to the other, becomes in Riemannian space the associate curvature of one congruence of curves with respect to another. The concept of distantal spread as a measure of the deviation from equidistance of the one family of curves with respect to the other suggests two corresponding spreads in Riemannian space: first, the distantal spread of a family of hypersurfaces with respect to a congruence of transversals; secondly, the distantal spread of one congruence of curves with respect to another. The latter concept is developed here only for the case when the curves of the two congruences lie on two-dimensional surfaces.

The paper deals with the relations between associate curvatures and distantal spreads and the conclusions which they imply concerning the connection between parallelism and equidistance.

It is assumed throughout that the linear element of the n -dimensional space V_n is defined by the positive definite quadratic form $ds^2 = g_{ij}dx^i dx^j$, the x 's being the coordinates of the V_n , and the g_{ij} 's real analytic functions of the x 's. The curves and surfaces considered are always assumed to be real and analytic.

1. *Associate curvature.* If C is a curve in V_n represented parametrically in terms of its arc by the equations $x^i = x^i(s)$, and λ^i are the components of a unit vector defined at each point of C , the associate curvature of the vector λ^i with respect to the curve C is

$$(1) \qquad 1/r = (g_{ij}\mu^i \mu^j)^{\frac{1}{2}}$$

* Graustein, "Parallelism and Equidistance in Classical Differential Geometry," *Transactions of the American Mathematical Society*, vol. 34 (1932), pp. 557-593.

where

$$(2) \quad \mu^i = (dx^k/ds)\lambda^i_{,k}$$

$\lambda^i_{,k}$ being the components of the covariant derivative of the vector λ^i with respect to the fundamental tensor g_{ij} .

The associate curvature $1/r$ is a measure of the deviation of the vectors λ^i from parallelism with respect to C in that its vanishing is the condition for parallelism.

The vector μ^i is known as the associate curvature vector of the vector λ^i with respect to the curve C ; it is orthogonal to the vector λ^i .

If the vectors λ^i are, in particular, tangent to C , μ^i and $1/r$ become respectively the principal normal vector and (first) curvature of C .

That the associate curvature of the vector λ^i with respect to C is based essentially on angle is evident from the fact that it can also be defined as

$$(3) \quad 1/r = \lim_{\Delta s \rightarrow 0} \Delta\theta/\Delta s$$

where $\Delta\theta$ is the angle between the vector λ^i at a point $P' : (s + \Delta s)$ of C and the direction at P' parallel with respect to C to the vector λ^i at $P : (s)$. The possibility of this definition was suggested by Bianchi* and carried through by Lipka† for the special case of the curvature of C .

Instead of the single curve C , we shall consider ordinarily n linearly independent congruences of curves C_k , ($k = 1, 2, \dots, n$), with unit tangent vectors $\lambda_k|{}^i = dx^i/ds_k$, s_k being the arc of the curves C_k . Then the associate curvature vector of the vectors $\lambda_k|{}^i$ (tangent to the curves C_k) with respect to the curves C_k is

$$(4) \quad \mu_{nk}|{}^j = \lambda_k|{}^i \lambda_{k|}{}^j{}_{,i}$$

We shall call this vector *the associate curvature vector of the curves C_k with respect to the curves C_k* , and its length

$$(5) \quad 1/r_{nk} = (g_{ij} \mu_{nk}|{}^i \mu_{nk}|{}^j)^{1/2}$$

the associate curvature of the curves C_k with respect to the curves C_k .

2. Equidistance of families of hypersurfaces with respect to congruences

* Bianchi, "Sul parallelismo vincolato di Levi-Civita nella metrica degli spazi curvi," *Rendiconti della Reale Accademia di Napoli*, ser. 3a, vol. 28 (1922), p. 161.

† Lipka, "Sulla curvatura geodetica delle linee appartenenti ad una varietà qualunque," *Rendiconti della Reale Accademia dei Lincei*, ser. 5, vol. 31 (1922), pp. 353-356.

of curves. Let $f_k(x^1, x^2, \dots, x^n) = c_k$, ($k = 1, 2, \dots, n$), be the equations of n independent families of hypersurfaces S_k in V_n . These hypersurfaces intersect in n linearly independent congruences of curves C_h , ($h = 1, 2, \dots, n$), of which $\lambda_h|^i = dx^i/ds_h$ are the unit tangent vectors. The hypersurfaces S_k of the family $f_k = c_k$ contain the $n - 1$ congruences C_h , ($h = 1, 2, \dots, n$; $h \neq k$), the curves C_k being transversals of these hypersurfaces.

Let d_k be the distance, measured along an arbitrary curve C_k , between a hypersurface $S_k : f_k = f_k^0$ and a neighboring hypersurface $S_k : f_k = f_k^0 + \Delta f_k$, and take the logarithmic directional derivative of d_k in the positive direction of a curve C_h lying in the hypersurface $S_k : f_k = f_k^0$. The limit of this derivative when Δf_k approaches zero is defined as the *distantial spread of the hypersurfaces S_k with respect to the curves C_k , measured along the curves C_h* :

$$(6) \quad 1/b_{hk} = \lim_{\Delta f_k \rightarrow 0} \partial \log d_k / \partial s_h \quad (h \neq k).$$

If s_k is the directed arc of an individual curve C_k ,

$$(7) \quad ds_k = (\partial f_k / \partial s_k)^{-1} df_k,$$

where $\partial f_k / \partial s_k$ is the directional derivative of f_k in the positive direction of the curves C_k . Hence, we obtain as the value of the foregoing limit

$$(8) \quad 1/b_{hk} = -\partial \log |\partial f_k / \partial s_k| / \partial s_h, \quad (h \neq k).$$

If $1/b_{hk} \equiv 0$, ($h = 1, 2, \dots, n$; $h \neq k$), then $\partial f_k / \partial s_k$ is a constant along each curve C_h , ($h \neq k$), and hence is a function of f_k . Then, ds_k , as given by (7) is an exact differential of a function $F(f_k)$, and $s_k = F(f_k)$ is the common directed arc, measured from a fixed hypersurface S_k , of all the curves C_k .

Noting that this argument may be reversed, we conclude that the *distantial spreads* $1/b_{hk}$ of the hypersurfaces S_k with respect to the curves C_k in the directions of the curves C_h , $h \neq k$, all vanish if and only if the hypersurfaces S_k are equidistant with respect to the curves C_k , in that each two of them cut segments of equal length from all the curves C_k . Thus the quantities $1/b_{hk}$ are actually measures of the deviation from equidistance of the hypersurfaces S_k with respect to the curves C_k .

Let θ_k denote the angle between the normal at a point P to the hypersurface S_k through P and the direction of the curve C_k through P . Since $\partial f_k / \partial s_k = (\Delta_1 f_k)^{1/2} \cos \theta_k$, where $(\Delta_1 f_k)^{1/2} = (g^{ij} f_{k|i} f_{k|j})^{1/2}$ is the magnitude of the gradient of f_k , we have

$$\begin{aligned} 1/b_{hk} &= -(\Delta_1 f_k)^{-1/2} \sec \theta_k \partial (f_{k|i} \lambda_h^i) / \partial s_h \\ &= -(\Delta_1 f_k)^{-1/2} \sec \theta_k (f_{k|i} \lambda_h^i)_{,j} \lambda_h^j \\ &= -(\Delta_1 f_k)^{-1/2} \sec \theta_k f_{k|i} \lambda_h^i \lambda_h^j + f_{k|i} \lambda_h^i \lambda_h^j. \end{aligned}$$

Now

$$f_{k|,ij}\lambda_{k|}^i\lambda_{h|}^j = \partial f_{k|,j}\lambda_{h|}^j / \partial s_k - f_{k|,j}\lambda_{h|}^j{}_{,i}\lambda_{k|}^i$$

since $f_{k|,ij} = f_{k|,ji}$. But $f_{k|,j}\lambda_{h|}^j = 0$, ($h = 1, 2, \dots, n$; $h \neq k$), since the vectors $\lambda_{h|}^j$ lie in the hypersurfaces S_k . Hence, using formula (4), we conclude that

$$(9) \quad \cos \theta_k / b_{hk} = (\Delta_1 f_k)^{-1/2} f_{k|,i} (\mu_{hk|}^i - \mu_{kh|}^i).$$

Since any two congruences of curves C_h and C_k lie in a two-dimensional surface, the vector $\mu_{hk|}^i - \mu_{kh|}^i$ at a point P lies in the surface containing the curves C_h and C_k through P ,* and hence can be written as a linear combination of the vectors $\lambda_{h|}^i$ and $\lambda_{k|}^i$.

$$(10) \quad \mu_{hk|}^i - \mu_{kh|}^i = a_h \lambda_{h|}^i + a_k \lambda_{k|}^i.$$

From (9) and (10) we conclude

THEOREM 1. *The hypersurfaces S_k are equidistant with respect to the curves C_k if and only if the vector which is the difference between the associate curvature vectors of the curves C_h and C_k with respect to one another lies always along the curves C_h for $h = 1, 2, \dots, n$; $h \neq k$.*

In particular, if all the congruences of curves C_h , ($h = 1, 2, \dots, n$; $h \neq k$), are parallel with respect to the curves C_k , and if the curves C_k are parallel with respect to all the congruences of curves C_h , then the hypersurfaces S_k are equidistant with respect to the curves C_k .

From (9) also follows

THEOREM 2. *A necessary and sufficient condition that the hypersurfaces of each family be equidistant with respect to the congruences of curves not contained in them is that the associate curvature vectors of each two congruences of curves with respect to each other be identical.*

As a corollary, it follows that if the curves of each congruence are parallel with respect to the curves of every other congruence, then the hypersurfaces of each family are equidistant with respect to the congruences of curves not contained in them.†

3. Distantal spread of one congruence of curves with respect to a second

* Struik, *Grundzüge der mehrdimensionalen Differentialgeometrie*, 1922, p. 53.

† This corollary has been proved directly by Corbellini, "Di una classe di varietà caratterizzate per mezzo del parallelismo," *Rendiconti della Reale Accademia dei Lincei*, ser. 6, vol. 4 (1926), p. 94.

congruence. Let us suppose, as before, that we have n independent congruences of curves C_h , ($h = 1, 2, \dots, n$), with unit tangent vectors $\lambda_h|_i$, which are the curves of intersection of n independent families of hypersurfaces S_k , $f_k(x^1, x^2, \dots, x^n) = c_k$, the curves C_k being taken as the transversals of the hypersurfaces S_k .

On each of the two-dimensional surfaces determined by the $n-2$ equations $f_i = c_i$, ($i = 1, 2, \dots, n$; $i \neq h, k$), there are, then, ∞^1 curves C_k , defined by the equation $f_k = c_k$, and ∞^1 curves C_h defined by the equation $f_h = c_h$. Considering the curves C_h and C_k on an arbitrary but fixed one of these surfaces, we form the logarithmic directional derivative in the positive direction of the curve C_h : $f_k = f_k^0$, of the distance, measured along an arbitrary curve C_k , between the curve C_h and a neighboring curve C_h : $f_k = f_k^0 + \Delta f_k$, and define the limit of this derivative when Δf_k approaches zero, as the *distantial spread of the curves C_h with respect to the curves C_k* .

It is readily shown that this distantial spread is given by the formula (8) and hence is identical with the distantial spread of the hypersurfaces S_k with respect to the curves C_k in the direction of the curves C_h . From (8) it follows that the curves C_h are equidistant with respect to the curves C_k in that on everyone of the two-dimensional surfaces in question each two curves C_h cut segments of equal length from the curves C_k if and only if the distantial spread of the curves C_h with respect to the curves C_k vanishes identically.

Since (9) is equivalent to (8), we have as the values of the distantial spreads of the curves C_h and C_k with respect to one another:

$$(11) \quad \begin{aligned} \cos \theta_k / b_{hk} &= (\Delta_1 f_k)^{-1/2} f_{k,i} (\mu_{hk}|^i - \mu_{kh}|^i) \\ \cos \theta_h / b_{kh} &= (\Delta_1 f_h)^{-1/2} f_{h,i} (\mu_{kh}|^i - \mu_{hk}|^i). \end{aligned}$$

Substituting the expression (10) in formulas (11), we find

$$a_h = -1/b_{kh}, \quad a_k = 1/b_{hk},$$

and hence

$$(12) \quad \mu_{hk}|^i - \mu_{kh}|^i = - (1/b_{kh}) \lambda_{h|i} + (1/b_{hk}) \lambda_{k|i}.$$

From this fundamental relation we obtain immediately the following theorems:

THEOREM 3. *If the curves of each of two congruences are parallel with respect to the curves of the other, then the curves of each congruence are equidistant with respect to the curves of the other.*

..

THEOREM 4. *If the curves of each of two congruences are equidistant*

with respect to the curves of the other, then their associate curvature vectors with respect to each other are identical.

THEOREM 5. *If the curves of each of two congruences are equidistant with respect to the curves of the other, and if the curves of one congruence are parallel with respect to the other, then the curves of the second are parallel with respect to the curves of the first.*

Using Theorem 4 and remembering that the associate curvature vector of the curves C_h with respect to the curves C_k is orthogonal to the curves C_h , we have the following result:

THEOREM 6. *If at each point P the associate curvature vectors of the curves of each of two congruences with respect to the curves of the other lie in the plane determined by the tangent vectors to the curves at P , then a necessary and sufficient condition that the curves of each congruence be parallel with respect to the curves of the other is that they be equidistant with respect to the other.*

4. *Case of an orthogonal system of hypersurfaces.* Let us consider a V_n admitting an n -tuply orthogonal system of hypersurfaces S_k , $f_k = c_k$, ($k = 1, 2, \dots, n$). The curves of intersection C_k form an ennuple of mutually orthogonal normal congruences.

Since $\mu_{kh}|^i \lambda_{k|i} = 0$, and since the directions of the curves C_k coincide with the normals to the hypersurfaces S_k , we have $f_{k|i} \mu_{kh}|^i = 0$. Formulas (11) then become

$$(13) \quad \begin{aligned} 1/b_{hk} &= (\Delta_1 f_k)^{-1/2} f_{k|i} \mu_{hk}|^i = \lambda_{k|i} \mu_{hk}|^i, \\ 1/b_{kh} &= (\Delta_1 f_h)^{-1/2} f_{h|i} \mu_{kh}|^i = \lambda_{h|i} \mu_{kh}|^i. \end{aligned}$$

Expressing the associate curvature vectors $\mu_{hk}|^i$ and $\mu_{kh}|^i$ in terms of the coefficients of rotation γ_{pqr} of the orthogonal ennuple of curves C_k , we have

$$\begin{aligned} \mu_{hk}|^i &= \lambda_{k|j} \lambda_{h|j}{}^i = \lambda_{k|j} \sum_{r,s} \gamma_{hrs} \lambda_r|{}^i \lambda_s|j = \sum_r \gamma_{hrk} \lambda_r|{}^i, \\ \mu_{kh}|^i &= \lambda_{h|j} \lambda_{k|j}{}^i = \lambda_{h|j} \sum_{r,s} \gamma_{krs} \lambda_r|{}^i \lambda_s|j = \sum_r \gamma_{krh} \lambda_r|{}^i, \end{aligned}$$

or

$$(14) \quad \begin{aligned} \mu_{hk}|^i &= - \sum_r \gamma_{rkh} \lambda_r|{}^i, \\ \mu_{kh}|^i &= - \sum_r \gamma_{rkh} \lambda_r|{}^i. \end{aligned}$$

Furthermore,

$$(15) \quad \dots \quad 1/r^2_{hk} = \sum_r \gamma^2_{rkh}, \quad 1/r^2_{kh} = \sum_r \gamma^2_{rkh}.$$

The formulas for the principal normal of the curves C_k ,

$$(16) \quad \mu_k|^i = -\sum_r \gamma_{rk} \lambda_r|^i;$$

and for the first curvature of these curves,

$$(17) \quad 1/\rho^2_k = \sum_r \gamma^2_{rk} \lambda_r|^i,$$

are special cases of formulas (14) and (15).

From (14) we have the following result:

THEOREM 7. *A necessary and sufficient condition that the curves C_h be parallel with respect to the curves C_k is that*

$$\gamma_{rhk} = 0, \quad (r = 1, 2, \dots, n).$$

As a special case of this theorem we have the condition that the curves C_k be geodesics: *

$$\gamma_{rk} = 0, \quad (r = 1, 2, \dots, n).$$

Formulas (14) through (17) and Theorem 7 hold for any orthogonal ennuple of congruences whatsoever. In this particular case, since the congruences of curves are all normal, the coefficients of rotation γ_{rhk} vanish for values of r, h , and k all distinct. Hence formulas (14) become

$$(18) \quad \mu_{hk}|^i = \gamma_{hkk} \lambda_k|^i, \quad \mu_{kh}|^i = \gamma_{khh} \lambda_h|^i.$$

Putting these values in equations (13) we have

$$(19) \quad 1/b_{hk} = \gamma_{hkk}, \quad 1/b_{kh} = \gamma_{khh}.$$

Since γ_{khh} (which is the negative of γ_{hkk}) is the orthogonal projection on C_h of the first curvature vector of C_k , we have

THEOREM 8. *The distantial spread at a point P of the curves C_h of a congruence with respect to the curves C_k of another is the negative of the orthogonal projection of the first curvature vector at P of the curve C_k on the tangent at P to the curve C_h .*

Formulas (18) can now be rewritten as

$$(20) \quad \mu_{hk}|^i = (1/b_{hk}) \lambda_k|^i, \quad \mu_{kh}|^i = (1/b_{kh}) \lambda_h|^i.$$

Hence we conclude the following theorems:

THEOREM 9. *The absolute value of the distantial spread of the curves of one congruence with respect to the curves of a second is equal to the associate*

* Eisenhart, *Riemannian Geometry*, p. 100.

curvature of the curves of the first congruence with respect to those of the second.

THEOREM 10. *If the curves of one congruence are equidistant with respect to the curves of another, then the curves of the first congruence are also parallel with respect to the curves of the second, and conversely.*

THEOREM 11. *The associate directions of the curves of one congruence with respect to the curves of a second coincide with the direction of the latter curves.*

From formula (18) it follows that

$$1/r^2_{hk} = \gamma^2_{hkk}, \quad (h \neq k).$$

Summing over h , omitting the value $h = k$, we have by formula (17)

$$\sum_h 1/r^2_{hk} = \sum_h \gamma^2_{hkk} = 1/\rho^2_k.$$

Hence

$$(21) \quad \sum_h 1/b^2_{hk} = \sum_h 1/r^2_{hk} = 1/\rho^2_k. \quad (h \neq k).$$

Considering this result from the point of view of the distastial spread of the family of hypersurfaces S_k with respect to their orthogonal trajectories C_k , we note that three properties are involved here:

- (a) The hypersurfaces S_k equidistant with respect to the curves C_k .
- (b) The curves C_k geodesic.
- (c) The curves C_h , ($h = 1, 2, \dots, n; h \neq k$), parallel with respect to the curves C_k .

From (21) we have

THEOREM 12. *If one of these three properties is valid, then so are the other two.**

5. *Oblique system of hypersurfaces.* We return to the general case when the n families of hypersurfaces are not orthogonal, and introduce the notation $1/\rho_{hk}$ for the projection of the associate curvature vector $\mu_{hk}|^i$ on $\lambda_k|^i$. Multiplying equation (12) first by $g_{ij}\lambda_k|^j$, and then by $g_{ij}\lambda_h|^j$, summing over i and j , we obtain the relations

$$(22) \quad \begin{aligned} 1/p_{hk} &= 1/b_{hk} - \cos \Omega_{hk}/b_{kh}, \\ 1/p_{kh} &= -\cos \Omega_{hk}/b_{hk} + 1/b_{kh}, \end{aligned}$$

* This result for the case when property (b) is given is not new. See Eisenhart, *loc. cit.*, p. 57. Such families of hypersurfaces are said to be geodesically parallel.

where Ω_{hk} is the angle between the curves C_h and C_k . Solving these two equations for $1/b_{hk}$ and $1/b_{kh}$, we find

$$(23) \quad \begin{aligned} \sin^2 \Omega_{hk}/b_{hk} &= 1/p_{hk} + \cos \Omega_{hk}/p_{kh}, \\ \sin^2 \Omega_{hk}/b_{kh} &= \cos \Omega_{hk}/p_{hk} + 1/p_{kh}. \end{aligned}$$

These results show that we have to deal with the following four properties of two congruences of curves:

- (A) $\left\{ \begin{array}{l} \text{Curves } C_h \text{ equidistant with respect to the curves } C_k: 1/b_{hk} = 0. \\ \text{Associate direction of the curves } C_h \text{ with respect to the curves } C_k \\ \text{orthogonal to the curves } C_k: 1/p_{hk} = 0. \end{array} \right.$
 $\left\{ \begin{array}{l} \text{Curves } C_k \text{ equidistant with respect to the curves } C_h: 1/b_{kh} = 0. \\ \text{Associate direction of the curves } C_k \text{ with respect to the curves } C_h \\ \text{orthogonal to the curves } C_h: 1/p_{kh} = 0. \end{array} \right.$

From equations (21) and (22) we conclude:

THEOREM 13. *Two non-orthogonal congruences of curves which have any two of the properties (A) have the other two also.*

The curves C_h and C_k may form an orthogonal system of curves without the families of hypersurfaces being an n -tuply orthogonal system. For such curves we have:

THEOREM 14. *Two orthogonal congruences of curves having one of the first two properties (A) and one of the last two have all four properties.*

Finally we have

THEOREM 15. *If the distantal spreads of the curves of each congruence with respect to the curves of the other are equals or negatives of each other, then so are the projections on the curves of each congruence of the associate curvature vectors, with respect to those curves, of the curves of the other congruence, and conversely.*

Using the quantities $1/p_{hk}$ and $1/p_{kh}$ we can express the components of the associate directions as follows:

$$(24) \quad \begin{aligned} \sin^2 \Omega_{hk} \mu_{hk|}^j &= -(\cos \Omega_{hk}/p_{hk}) \lambda_{h|}^j + (1/p_{hk}) \lambda_{k|}^j + v_{hk|}^j, \\ \sin^2 \Omega_{hk} \mu_{kh|}^j &= (1/p_{kh}) \lambda_{h|}^j - (\cos \Omega_{hk}/p_{kh}) \lambda_{k|}^j + v_{kh|}^j, \end{aligned}$$

where $v_{hk|}^j = v_{kh|}^j$ is the common component of the associate directions orthogonal to the plane of $\lambda_{h|}^j$ and $\lambda_{k|}^j$

ON A RATIONAL SURFACE OF ORDER 12 IN 9-SPACE AND ITS PROJECTIONS.

By B. C. WONG.

1. Introduction. The quadratic transformation *

$$(1) \quad \rho x_{ij} = x_i x_j \quad [i, j = 1, 2, 3, 4]$$

where x_1, x_2, x_3, x_4 are the homogeneous coördinates of a point of a 3-space S_3 and x_{11}, \dots, x_{34} are those of a 9-space S_9 , transforms the points of S_3 into the points of an octavic variety V_3^8 of order 8 and dimension 3 in S_9 . To a plane of S_3 corresponds a Veronese quartic surface † lying on V_3^8 and to a quadric surface of S_3 corresponds a section of V_3^8 by a hyperplane of S_9 . Of interest is the surface Φ^{12} of order 12 on V_3^8 which corresponds to a general cubic surface F^3 of S_3 . It has 27 conics whose relative positions can best be studied by reference to the relative positions of the 27 lines on F^3 . By representing F^3 upon a plane ϕ by means of the ∞^3 cubic curves through six general points P_1, P_2, \dots, P_6 of ϕ , we see that Φ^{12} can be represented upon ϕ by means of the ∞^9 sextic curves with six nodes at P_λ [$\lambda = 1, 2, \dots, 6$].

If we let $(y_1 : y_2 : y_3)$ be the coördinates of a point of ϕ and $(p_1^{(\lambda)} : p_2^{(\lambda)} : p_3^{(\lambda)})$ be those of the six points P_λ , we may take the equations

$$f_i(y_1 : y_2 : y_3) \equiv \sum_1^8 u_{iklm} y_k y_l y_m = 0 \quad [i = 1, 2, 3, 4]$$

where

$$f_i(p_1^{(\lambda)} : p_2^{(\lambda)} : p_3^{(\lambda)}) \equiv 0$$

to be the equations of four general cubics of the ∞^3 -system $|c^3|$ of cubics passing through the six points P_λ in ϕ . Then the equations

$$(2) \quad x_1 : x_2 : x_3 : x_4 = f_1 : f_2 : f_3 : f_4$$

are those of the cubic transformation which transforms a point of ϕ into a point of F^3 . They are the parametric equations of F^3 .

* This transformation was mentioned in B. C. Wong, "A hypersurface of order $2r-1$ in r -space," *American Journal of Mathematics*, vol. 54 (1932), pp. 293-298. It was used by W. G. Welchman in his paper, "On elliptic quartic curves with assigned points and chords," and by J. A. Todd in his paper, "Some enumerative results for elliptic quartic curves," *Proceedings of the Cambridge Philosophical Society*, vol. 27 (1931), pp. 20-23 and 538-542 respectively, to solve certain problems on elliptic quartic curves.

† Bertini, *Projective Geometrie Mehrdimensionaler Räume* (1924), Chapter 16.

Now the equation of a general sextic curve of the ∞^3 -system $|c^6|$ of sextic curves having the points P_λ for nodes is of the form

$$\sum_1^4 A_{ij} f_i f_j = 0.$$

This sextic c^6 has for image in S_3 the sextic curve C^6 in which F^3 is met by the quadric surface

$$\sum_1^4 A_{ij} x_i x_j = 0.$$

Now corresponding to C^6 is the curve Γ^{12} of order 12 which is the intersection of Φ^{12} and the hyperplane

$$\sum_1^4 A_{ij} x_{ij} = 0.$$

Then Γ^{12} is the image of c^6 in ϕ under the correspondence, whose equations are

$$\sigma x_{ij} = f_i f_j.$$

These are also the parametric equations of Φ^{12} .

If we now project ϕ^{12} from a general 5-space of S_5 upon a 3-space, in particular the 3-space S_3 containing F^3 , we have for projection a surface Φ'^{12} which is represented upon ϕ by the sextic curves of a certain ∞^3 -system $|c^6|'$ of $|c^6|$. If the center of projection is in a particular position relative to Φ^{12} , the projection will possess particular properties or will be of lower order, say $12 - \nu$. The properties of these different projections can be described without difficulty. The question arises as to how to choose the center of projection so that the curves of $|c^6|'$ will be the Jacobians of the nets of cubics of $|c^3|$. If the selection is properly made, the surface Φ'^{12} will be reciprocal to F^3 . The sextic curves C^6 , which are the images of the curves of $|c^6|'$ under the transformation (2) and which have for images under the transformation (1) the sections of Φ^{12} by the hyperplanes of S_5 through the center of projection, will be the intersections of F^3 and the polar quadric surfaces of all the points of S_3 with respect to F^3 .

In what follows we shall, after describing very briefly some of the various projections of Φ^{12} , determine the 5-space S_5 for center of projection so that the projection Φ'^{12} will be reciprocal to F^3 . Then, as a special case, we shall obtain the Steiner's quartic surface from the Φ^{12} corresponding to the four-nodal cubic surface in S_3 . Finally, we shall obtain from the same Φ^{12} a surface reciprocal to a certain nonic surface whose equation is

$$x_1^{1/3} + x_2^{1/3} + x_3^{1/3} + x_4^{1/3} = 0.$$

2. *Properties of $\Phi^{12-\nu}$.* If the 5-space S_5 , the center of projection, has ν [$0 \leq \nu \leq 6$] general points in common with the Φ^{12} corresponding to a general cubic surface F^3 of S_5 , the projection of Φ^{12} is a surface $\Phi^{12-\nu}$ of order $12 - \nu$. It has a double curve of order $b = (\nu^2 - 21\nu + 102)/2$, $j = 42 - 4\nu$ pinch points and $t = (10 - \nu)(9 - \nu)(8 - \nu)/6 + 3(\nu - 7)$ triple points. It is of class $m = 33$ for all permissible values of ν . If $\nu = 0$, then, for Φ^{12} , $b = 51$, $j = 42$, and $t = 99$. Now let S_5 be incident with μ planes, each containing a conic of Φ^{12} , and keeping $\nu = 0$ we see that the double curve is composed of a curve of order $51 - \mu$ and μ lines on each of which are two of the 42 pinch points.

Now let S_5 contain a conic of Φ^{12} and have ν' [$0 \leq \nu' \leq 3$] general points in common with Φ^{12} , and we have for projection a surface $\Phi^{7-\nu'}$ of order $7 - \nu'$ and class 20 on which lies a double curve of order $b = (\nu'^2 - 11\nu' + 26)/2$. There are $j = 4(4 - \nu')$ pinch points and $t = (2 - \nu')(3 - \nu')(7 - \nu')/6$ triple points. For $\nu' = 3$, we have the familiar rational quartic surface with a double line.* If S_5 contains two non-intersecting conics of Φ^{12} , then the projection is a quadric surface; but if S_5 contains two intersecting conics, the projection is a quartic surface with a double conic.† Finally, if S_5 contains an elliptic sextic curve of Φ^{12} or contains three conics of Φ^{12} which are incident in pairs, the projection is a general cubic surface.

The results above can be verified easily by considering the representation of each of the projections upon the plane ϕ or by means of known formulas concerning rational surfaces.‡

3. *The determination of S_5 so that Φ^{12} will be reciprocal to F^3 .* Denote the 27 conics on Φ^{12} by the symbols $\alpha_1, \dots, \alpha_6, \beta_1, \dots, \beta_6, \gamma_{12}, \gamma_{13}, \dots, \gamma_{56}$ in conformity with the double-six notation $a_1, \dots, a_6, b_1, \dots, b_6, c_{12}, c_{13}, \dots, c_{56}$ for the 27 lines on F^3 . The 27 conics are such that each intersects 10 others each in one point and that they all form 45 triples, the members of each triple being incident in pairs. The planes of non-incident conics are skew.

Consider, in particular, the conic γ_{ij} . It intersects γ_{kl} in the point $\gamma_{ij} \cdot \gamma_{kl}$ and γ_{mn} in the point $\gamma_{ij} \cdot \gamma_{mn}$ [$i \neq j \neq k \neq l \neq m \neq n$] while γ_{kl} intersects γ_{mn} in the point $\gamma_{kl} \cdot \gamma_{mn}$. Denote the plane of these three points by $(\gamma_{ij} \cdot \gamma_{kl} \cdot \gamma_{mn})$. By permuting the subscripts we obtain 15 such planes. Now γ_{ij} intersects the conic α_i in a point $\gamma_{ij} \cdot \alpha_i$ and the conic β_j in a point

* Jessop, *Quartic Surfaces*, Cambridge Press, 1916, Chapter VI.

† *Ibid.*, Chapter IV.

‡ Cayley, *Collected Mathematical Papers*, vol. 8, pp. 388-393.

$\gamma_{ij} \cdot \beta_j$ while α_i and β_j intersect in a point $\alpha_i \cdot \beta_j$. Denote the planes of these three points by $(\gamma_{ij} \cdot \alpha_i \cdot \beta_j)$. Thus, we get 30 planes of this type.

Now the points on each of the 27 conics corresponding to the points on each of the 27 lines of F^3 , where the planes through the line are doubly tangent to F^3 , are paired in involution. To the involution on γ_{ij} belong the three pairs of points $\gamma_{ij} \cdot \gamma_{kl}$, $\gamma_{ij} \cdot \gamma_{mn}$; $\gamma_{ij} \cdot \gamma_{km}$, $\gamma_{ij} \cdot \gamma_{ln}$; $\gamma_{ij} \cdot \gamma_{kn}$, $\gamma_{ij} \cdot \gamma_{lm}$ and also the two pairs $\gamma_{ij} \cdot \alpha_i$, $\gamma_{ij} \cdot \beta_j$; $\gamma_{ij} \cdot \alpha_j$, $\gamma_{ij} \cdot \beta_i$. The lines determined by all these pairs are all concurrent at a point C_{ij} , the center of involution. In a similar manner we obtain the centers C_{kl} and C_{mn} of involution for the conics γ_{kl} and γ_{mn} respectively. These three points C_{ij} , C_{kl} , C_{mn} lie on a line which may be denoted by $C_{ij} \cdot C_{kl} \cdot C_{mn}$ and this line lies in the plane $(\gamma_{ij} \cdot \gamma_{kl} \cdot \gamma_{mn})$. There are 15 such lines. Let the centers of involution on the conics α_i , β_j be denoted by A_i , B_j respectively. The three points C_{ij} , A_i , B_j lie on a line $C_{ij} \cdot A_i \cdot B_j$ of the plane $(\gamma_{ij} \cdot \alpha_i \cdot \beta_j)$. There are 30 such lines.

The 27 points A_i , B_j , C_{ij} and the 45 lines just obtained form a configuration such that each point is on five lines and each line contains three points. Now we wish to show that this configuration lies completely in a 5-space. Consider two skew lines, say $C_{12} \cdot C_{34} \cdot C_{56}$ and $C_{13} \cdot C_{25} \cdot C_{46}$, of the configuration. They determine four other lines, namely, $C_{12} \cdot C_{46} \cdot C_{35}$, $C_{13} \cdot C_{56} \cdot C_{24}$, $C_{34} \cdot C_{25} \cdot C_{16}$, $C_{24} \cdot C_{35} \cdot C_{16}$. All these six lines lie in a 3-space S_3 . Now choose a line through one of the nine points already in S_3 , say the line $C_{12} \cdot C_{26} \cdot C_{45}$ which meets S_3 in the point C_{12} . This line determines with S_3 a 4-space S_4 . Now it is easy to see that this S_4 contains all the 15 lines of the type $C_{ij} \cdot C_{kl} \cdot C_{mn}$. But the line $C_{12} \cdot A_1 \cdot B_2$, having the point C_{12} in common with S_4 , determines with it a 5-space S_5 . In this S_5 are obviously the lines $C_{13} \cdot A_1 \cdot B_3$, $C_{13} \cdot A_3 \cdot B_1$, $C_{14} \cdot A_1 \cdot B_4$, and the rest of the 30 lines of the type $C_{ij} \cdot A_i \cdot B_j$.

This is the very S_5 we have been seeking. Taking it as the center of projection, we see that the 27 double lines on the projection Φ^{12} come from the 27 conics of Φ^{12} and that the 45 triple points come from the 45 triples of points on Φ^{12} of which 15 are of the type $\gamma_{ij} \cdot \gamma_{kl}$, $\gamma_{ij} \cdot \gamma_{mn}$, $\gamma_{kl} \cdot \gamma_{mn}$ and 30 are of the type $\gamma_{ij} \cdot \alpha_i$, $\gamma_{ij} \cdot \beta_j$, $\alpha_i \cdot \beta_j$. There is no need of describing the properties of Φ^{12} as they are already familiar.*

4. *The Steiner's quartic surface.* As a very special case we consider the Φ^{12} in S_9 corresponding to a four-nodal cubic surface F^3 . Let F^3 have the equation

* Salmon-Rogers, *Analytic Geometry of Three Dimensions*, 5th ed., vol. 2, Chapter XVIIa.

$$1/x_1 + 1/x_2 + 1/x_3 + 1/x_4 = 0$$

or the equations

$$x_1 : x_2 : x_3 : x_4 = 1/y_1 : 1/y_2 : 1/y_3 : 1/y_4$$

where

$$y_1 + y_2 + y_3 + y_4 = 0.$$

The base points of the cubic curves of representation in ϕ are the six vertices of the quadrilateral y_1, y_2, y_3, y_4 . Regarding

$$(3) \quad \begin{aligned} f_1 &\equiv y_2 y_3 y_4 = 0, \\ f_2 &\equiv y_3 y_4 y_1 = 0, \\ f_3 &\equiv y_4 y_1 y_2 = 0, \\ f_4 &\equiv y_1 y_2 y_3 = 0 \end{aligned}$$

as four independent cubics of the ∞^3 -system $|c^3|$, we have

$$(4) \quad \begin{aligned} \Sigma A_{ij} f_i f_j &= A_{11} y_2^2 y_3^2 y_4^2 + A_{22} y_3^2 y_4^2 y_1^2 + A_{33} y_4^2 y_1^2 y_2^2 + A_{44} y_1^2 y_2^2 y_3^2 \\ &\quad + 2y_1 y_2 y_3 y_4 (A_{12} y_3 y_4 + A_{13} y_4 y_2 + A_{23} y_1 y_4 \\ &\quad + A_{34} y_1 y_2 + A_{24} y_1 y_3 + A_{14} y_2 y_3) = 0 \end{aligned}$$

for the equation of the ∞^9 -system $|c^6|$ of sextics having nodes at the base points of $|c^3|$. The equation of Φ^{12} may be written

$$\sigma x_{ij} = 1/y_i y_j.$$

Now it is easy to see that Φ^{12} has four conical points given by

$$x_{ii} = 1, \quad x_{ij} = x_{kj} = x_{kl} = 0 \quad [i, j = 1, 2, 3, 4].$$

Projecting Φ^{12} from any 5-space containing these four conical points upon S_3 , we obtain a Steiner's quartic surface. But we wish to project from a particular 5-space so that the projection will be the exact reciprocal of F^3 , that is, will have the equation

$$x_1^{1/2} + x_2^{1/2} + x_3^{1/2} + x_4^{1/2} = 0.$$

There are nine conics on Φ^{12} of which six intersect three by three on the conical points and the remaining three are incident in pairs. Denote the first six conics by $\beta_{12}, \beta_{13}, \dots, \beta_{34}$ where β_{ij} passes through the conical points

$$X_{ii}(x_{ii} = 1, x_{ij} = x_{kj} = x_{kl} = 0) \quad \text{and} \quad X_{jj}(x_{jj} = 1, x_{ij} = x_{kj} = x_{kl} = 0)$$

and the other three by $\gamma_{12,34}, \gamma_{13,42}, \gamma_{14,23}$ where $\gamma_{ij,kl}$ is incident with β_{ij}

and β_{kl} . The points on each of these three conics, corresponding to the points on each of the three coplanar lines of F^3 not passing through the nodes where the planes through the line are doubly tangent to F^3 , are paired in involution. Let $C_{12,34}$, $C_{13,42}$, $C_{14,23}$ be the centers of involutions on the conics $\gamma_{12,34}$, $\gamma_{13,42}$, $\gamma_{14,23}$ respectively. They have the coördinates, as a little computation will show,

$$\begin{aligned} & x_{11} : x_{22} : x_{33} : x_{44} : x_{12} : x_{34} : x_{13} : x_{42} : x_{14} : x_{23} \\ C_{12,34} & (0 : 0 : 0 : 0 : 0 : 0 : 1 : 1 : -1 : -1), \\ C_{13,42} & (0 : 0 : 0 : 0 : 1 : 1 : 0 : 0 : -1 : -1), \\ C_{14,23} & (0 : 0 : 0 : 0 : 1 : 1 : -1 : -1 : 0 : 0). \end{aligned}$$

These three points evidently lie on a line. Now the 5-space containing the four conical points and this line is the S_5 we desire to have for center of projection. Every hyperplane through this S_5 has an equation of the form

$$(5) \quad A_{12}x_{12} + A_{34}x_{34} + A_{13}x_{13} + A_{42}x_{42} + A_{14}x_{14} + A_{23}x_{23} = 0$$

where

$$A_{12} + A_{34} = A_{13} + A_{42} = A_{14} + A_{23}.$$

Now let

$$\begin{aligned} A_{12} : A_{34} : A_{13} : A_{42} : A_{14} : A_{23} \\ = u_3 + u_4 : u_1 + u_2 : u_4 + u_2 : u_1 + u_3 : u_2 + u_3 : u_1 + u_4 \end{aligned}$$

and equation (5) becomes

$$\begin{aligned} (u_3 + u_4)x_{12} + (u_1 + u_2)x_{34} + (u_4 + u_2)x_{13} + (u_1 + u_3)x_{42} \\ + (u_2 + u_3)x_{14} + (u_1 + u_4)x_{23} = 0. \end{aligned}$$

Putting these values of the coefficients and also $A_{ii} = 0$ in equation (4), we have, after reduction,

$$u_1y_1^2 + u_2y_2^2 + u_3y_3^2 + u_4y_4^2 = 0$$

for the equation of the Jacobians of the nets of the ∞^3 -system $|c^3|$ of cubic curves representing the four-nodal cubic surface F^3 . These Jacobian conics represent the plane sections of the Steiner's quartic surface whose equation is evidently

$$x_1^{\frac{1}{2}} + x_2^{\frac{1}{2}} + x_3^{\frac{1}{2}} + x_4^{\frac{1}{2}} = 0.$$

5. *The Φ^{12} reciprocal to a certain nonic surface.* The totality of cubic curves in the plane ϕ may be taken to represent the normal rational surface F^9 in S_9 . Projecting F^9 from a certain S_5 upon S_3 , we obtain a surface F^9

which may be represented upon ϕ by the system of cubic curves given by the equation

$$A_1y_1^3 + A_2y_2^3 + A_3y_3^3 + A_4y_4^3 = 0,$$

where

$$y_1 + y_2 + y_3 + y_4 = 0,$$

and therefore has the equation

$$x_1^{1/3} + x_2^{1/3} + x_3^{1/3} + x_4^{1/3} = 0.$$

Consider again the Φ^{12} corresponding to the four-nodal cubic surface F^3 . In order to obtain a Steiner's quartic surface for projection we took for center of projection an S_5 containing the four conical points of Φ^{12} . Now suppose we take for center of projection the S_5 determined by the six vertices other than X_{ii} of the coördinate simplex given by the coördinates

$$x_{ii} = 0, \quad x_{ij} = 1, \quad x_{ik} = x_{jk} = x_{ki} = 0.$$

The ∞^3 hyperplanes through this S_5 have for equation

$$A_{11}x_{11} + A_{22}x_{22} + A_{33}x_{33} + A_{44}x_{44} = 0$$

and the Jacobian sextic curves of the nets of the ∞^3 -system of cubic curves given by (6) have for equation

$$A_{11}/y_1^2 + A_{22}/y_2^2 + A_{33}/y_3^2 + A_{44}/y_4^2 = 0.$$

These sextics are the curves of representation upon ϕ of the projection Φ'^{12} of Φ^{12} reciprocal to F'^9 . The equation of Φ'^{12} is evidently

$$x_1^{-1/2} + x_2^{-1/2} + x_3^{-1/2} + x_4^{-1/2} = 0.$$

The properties of F'^9 have been studied recently by A. R. Williams* and those of Φ'^{12} can be easily deduced either by dualizing F'^9 or from its representation upon ϕ .

UNIVERSITY OF CALIFORNIA.

* "Analogues of the Steiner surface and their double curves," *Bulletin of the American Mathematical Society*, vol. 39 (1933), pp. 621-626.

REDUCIBLE EXCEPTIONAL CURVES OF THE FIRST KIND.*

By S. F. BARBER † and OSCAR ZARISKI

Introduction. In a birational transformation between two surfaces to certain curves of one of the surfaces may correspond *simple* points of the other. Such curves have been called *exceptional curves*. Exceptional curves have been subdivided into those of the first or second kind according as a point of the curve is not or is transformed into a curve of the other surface. The ones previously considered have either been irreducible or consisted of at most two components. A treatment of reducible exceptional curves with components all simple is found in (⁴, Chap. II, 6). We shall consider exceptional curves of the first kind with s irreducible components, each counted a certain number of times. This case arises when the birational transformation possesses infinitely near fundamental points on one or several irreducible algebroid branches.

In the sequel we shall use some fundamental notions of linear systems of curves on an algebraic surface and the theory of singularities as developed by Enriques (³; pp. 327-399).

Sections 1 and 2 of this paper are devoted to definitions and the derivation of the virtual degree and genus of an exceptional curve of the first kind; the presentation here is parallel to but more complete than that found in (⁴, Chap. II, 6). In sections 3 and 4 we introduce s fundamental points of the birational transformation, mentioned above, on that surface where the given exceptional curve has been changed into a point, and limit ourselves to the case in which these fundamental points lie on a single irreducible algebroid branch. In 5 we consider the correspondence between the irreducible components of the given curve on the one surface and the immediate neighborhood of each of the fundamental points on the other surface. In 6 we determine the multiplicities of these fundamental points on branches of lowest order passing through the first α ($\alpha \leq s$) of them. In 7 we describe the intersections of the irreducible components of the given exceptional curve, using extensively the notion of proximate points introduced by Enriques in his theory of singularities (³, p. 381); in 8 we study the classification of the s fundamental points as free points or satellites when the intersection numbers

* Presented to the Society, March 30, 1934.

† National Research Fellow.

of the components of the exceptional curve are known. In 9 we extend to reducible exceptional curves of the first kind the characterization of irreducible exceptional curves of the first kind as found in (1). Finally in 10 we consider the situation when the fundamental points lie on several irreducible algebroid branches.

1. *Definitions.* Let F be an algebraic surface in S_r , free from singularities, and let Σ be a linear system of curves C on F , of dimension r . We shall say for brevity that a curve L on F , irreducible or not, is a total fundamental curve of Σ , if L is a fixed component of a linear subsystem, ∞^{r-1} , of Σ and if the curves of this subsystem do not have other fixed components outside L . We shall adopt the following

Definition. (4, Chap. II, 6). A curve L on F is an exceptional curve if there exists on F a linear system Σ of curves C such that

- (i) Σ is irreducible, simple, of dimension $r \geq 3$;
- (ii) L is a total fundamental curve of Σ ;
- (iii) The ∞^{r-1} curves C_1 , which together with L give total curves of Σ , form a linear system Σ_1 of effective degree one less than the effective degree of Σ .

We recall that the effective degree of a linear system is the number of variable intersections of two curves of the system. If the system is of dimension r , its effective degree is $\geq r - 1$. It follows that the system Σ_1 is necessarily irreducible, since by (ii) it is free from fixed components and by (i) ($r \geq 3$) and (iii) it is of positive effective degree and therefore cannot be composed of the curves of a pencil.

If it is possible to find the above system Σ in such a manner that it should have no base points on L , then L is called an exceptional curve of the *first kind*. If this is not possible, L is called an exceptional curve of the *second kind*.

If we refer the ∞^r curves C of Σ to the hyperplanes of an S_r , we obtain in S_r a surface \bar{F} birationally equivalent to F , since Σ is simple. This surface is of order n , where n is the effective degree of Σ , and its hyperplane sections \bar{C} correspond to the curves C of Σ . In this correspondence to the subsystem $\Sigma_1 + L$ of Σ there corresponds the system of hyperplane sections of \bar{F} on a fixed point O of S_r . This point O is on \bar{F} , since the effective degree of Σ_1 is less than the effective degree of Σ , and is a *simple* point of \bar{F} , since by (iii) two hyperplane sections of \bar{F} on O have $n - 1$ variable intersections. Thus in the birational correspondence T between F and \bar{F} the curve L (irreducible or not) is transformed into a *simple* point O of \bar{F} (a fundamental point of T^{-1}) and this is in agreement with the usual definition of an exceptional curve.

It is one of the purposes of this paper to study the nature of the correspondence between L and the neighborhood of O when L is reducible.

The fundamental points of T (on F) are at the base points of Σ . Hence an exceptional curve L is of the first or of the second kind (see above definition) according as it is or is not possible to transform L into a simple point of a surface \bar{F} , birationally equivalent to F , in such a manner that no point of L is at the same time transformed into a curve of \bar{F} .

In the sequel we shall be concerned only with exceptional curves of the first kind. We observe, however, that the properties of exceptional curves of the second kind can be deduced from those of the first kind. In fact, it is always possible to transform birationally F into a surface \bar{F} in such a manner that the transformation should possess no fundamental points on \bar{F} and that to Σ there should correspond on \bar{F} a linear system $\bar{\Sigma}$, free from base points. For this it is sufficient to refer the sections of F by hyperquadrics on a base point of Σ to the hyperplanes of a linear space and to apply this procedure successively to the transformed surfaces until all the base points of Σ disappear. To the exceptional curve L there will correspond on \bar{F} an exceptional curve \bar{L} of the *first kind*. However, in general \bar{L} will consist of the transform proper of L and of the fundamental curves of the transformation on \bar{F} , arising from the base points of Σ .

2. *The virtual characters of an exceptional curve of the first kind.*
Theorem 1 (⁴, Chap. II, 6). *If L is an exceptional curve of the first kind, then $(L^2) = -1$ and $[L] = 0$.**

Proof. Let H denote the set of effective base points of Σ . We have $(C \cdot L) = ((C_1 + L) \cdot L) = 0$, since L , a total fundamental curve of Σ , does not have variable intersections with a \bar{C} and since no base point of Σ is on L . Consequently $(C_1 \cdot L) = -(L^2)$. Now $C_1 + L$ is the limit of an irreducible curve C and hence C_1 and L must have at least one point in common (Principle of degeneration of Enriques²). It follows that $(C_1 \cdot L) > 0$ and hence $(L^2) < 0$. We also have $\dagger n = (C^2)_H = (C_1^2)_H + 2(C_1 \cdot L) + (L^2) = (C_1^2)_H - (L^2)$. From the hypothesis that the effective degree of Σ_1 is $n - 1$, it follows that $(C_1^2)_H \geq n - 1, \dagger$ and that consequently $(L^2) \geq -1$.

* (L^2) is the virtual degree, $[L]$ the virtual genus. In evaluating these characters of L , we consider L as *virtually free from base points*.

$\dagger (C^2)_H$ denotes the virtual degree of C with respect to the assigned set H of base points, in this case the effective degree of Σ .

\ddagger The set of effective base points of Σ_1 certainly includes H , since $C_1 + \bar{L} = C$. *A priori* we cannot exclude the possibility that Σ_1 may possess accidental base points, whence the necessity of using the inequality sign.

It follows that $(L^2) = -1$ and incidentally

$$(1) \quad (C_1^2)_H = n - 1;$$

$$(2) \quad (C_1 \cdot L) = 1.$$

We observe that from (1) it follows that H is also the set of effective base points of Σ_1 and hence $[C_1]_H$ gives the effective genus of C_1 . Since the curves C_1 correspond to the hyperplane sections of \bar{F} on a simple point O , it follows that $[C_1]_H = [C]_H$. Furthermore

$$[C]_H = [C_1]_H + [L] + (C_1 \cdot L) - 1 = [C_1]_H + [L].$$

Hence $[L] = 0$, q. e. d.

The exceptional curve L may be reducible and some of its irreducible components may occur in L to a certain multiplicity. Let then $L = k_1 L_1 + k_2 L_2 + \cdots + k_s L_s$, where L_1, L_2, \cdots, L_s are the distinct irreducible components of L and where k_1, k_2, \cdots, k_s are positive integers. From (2) it follows that C_1 intersects one and only one of the components L_i and that this must be a simple component of L . Choosing our notation properly we may assume that

$$(3) \quad (C_1 \cdot L_1) = 1, \quad (C_1 \cdot L_i) = 0 \quad (i = 2, \cdots, s),$$

$$(3') \quad k_1 = 1.$$

Since $(C \cdot L_i) = 0$ ($i = 1, \cdots, s$), it follows that

$$(4) \quad (L \cdot L_1) = -1, \quad (L \cdot L_i) = 0 \quad (i = 2, \cdots, s),$$

all curves L_i being considered as virtually free from base points. The relations (4) can be rewritten as follows:

$$(5) \quad \begin{aligned} (L_1^2) + \sum_{i=2}^s k_i (L_1 \cdot L_i) &= -1, \\ (L_j \cdot L_1) + \sum_{i=2}^s k_i (L_j \cdot L_i) &= 0 \quad (j = 2, \cdots, s). \end{aligned}$$

We observe that the curve $L_1 + L_2 + \cdots + L_s$ must be *connected*, i. e. it cannot be represented as the sum of two curves having no points in common. This follows immediately from the fact that C_1 meets only the curve L_1 and that $C_1 + L$ is the limit of the irreducible curve C . Since $(L_i \cdot L_j) \geq 0$, if $i \neq j$, and since, by the preceding remark, for a given value of i one at least of the intersection numbers $(L_i \cdot L_j)$, $i \neq j$, must be positive, it follows from (5) that the virtual degree (L_i^2) of any component L_i of L is negative.

Applying the well known formula for the virtual genus of a reducible curve, we find

$$[L] = \sum_{i=1}^s k_i p_i + \sum_{i=1}^s [k_i(k_i - 1)/2] (L_i)^2 \\ + (1/2) \sum_{\substack{i=1 \\ i \neq j}}^s \sum_{j=1}^s k_i k_j (L_i \cdot L_j) - \sum_{i=1}^s k_i + 1,$$

where $p_i = [L_i]$. By means of (4) or the equivalent relations (5) this expression of $[L]$ can be simplified:

$$[L] = \sum_{i=1}^s k_i p_i - k_1/2 - (1/2) \sum_{i=1}^s k_i (L_i)^2 - \sum_{i=1}^s k_i + 1.$$

Since by Theorem 1 we have $[L] = 0$ and since $k_1 = 1$ we arrive at the following relation:

$$(6) \quad \sum_{i=1}^s k_i [2p_i - (L_i)^2 - 2] + 1 = 0.$$

3. *The fundamental points in the first neighborhood of O .* We now proceed to study the system Σ_1 , the residual of Σ with respect to L . By (3) the curves L_2, \dots, L_s are fundamental curves of Σ_1 . On the contrary L_1 is not a fundamental curve of Σ_1 , since $(C_1 \cdot L_1) = 1$, and since the intersection of C_1 and L_1 cannot be a fixed intersection, because by (1) the effective base points of Σ_1 coincide with those of Σ and are all outside L . Let us assume for the sake of generality that $L_2 + \dots + L_s$ consists of several connected curves $\bar{L}^{(1)}, \bar{L}^{(2)}, \dots, \bar{L}^{(\sigma)}$, having two by two no points in common (maximal connected components of $L_2 + \dots + L_s$). Each $\bar{L}^{(i)}$ is then itself a fundamental curve of Σ_1 . Let us consider for instance $\bar{L}^{(1)}$ and let us denote by Σ_2 the residual system of Σ_1 with respect to $\bar{L}^{(1)}$. Let C_2 denote a generic curve of Σ_2 . What are the possible fixed components of Σ_2 ? L_1 is not a fixed component since it is not fundamental for Σ_1 . Neither can any component of $\bar{L}^{(i)}$, $i > 1$, be a fixed component of Σ_2 . In fact we have $(C_1 \cdot L_1) = ((C_2 + \bar{L}^{(1)}) \cdot L_1) = 1$, and since $(\bar{L}^{(1)} \cdot L_1) > 0$ (the curve $L_1 + \dots + L_s$ is connected), and $(C_2 \cdot L_1)$ cannot be negative, it follows that $(C_2 \cdot L_1) = 0$ and incidentally that

$$(7) \quad (\bar{L}^{(1)} \cdot L_1) = 1.$$

If any component of $\bar{L}^{(i)}$, $i > 1$, were a fixed component of Σ_2 , then the whole curve $\bar{L}^{(i)}$ would be a fixed component of Σ_2 , since it is a fundamental curve of Σ_1 . We would have then $C_2 = \bar{L}^{(i)} + \bar{C}_2$ and $(\bar{C}_2 \cdot L_1) = -(\bar{L}^{(i)} \cdot L_1) \leq 0$, and this is impossible since L_1 is not a fixed component of Σ_2 .

Thus the only possible fixed components of Σ_2 are either the irreducible

components of $\bar{L}^{(1)}$ or irreducible curves which are not components of L . We can, however, avoid this second possibility by replacing the system Σ_1 by the system $\Sigma^{(1)}$, which corresponds on F to the system cut out on \bar{F} by the hyperquadrics (instead of by the hyperplanes) passing through the fundamental point O . This system $\Sigma^{(1)}$ contains all the reducible curves $C + C_1$, and it is obvious that the preceding conclusions apply as well to $\Sigma^{(1)}$ as to Σ_1 . The curves of $\Sigma^{(1)}$ which contain $\bar{L}^{(1)}$ as a fixed component do not automatically possess other fixed components which are not components of L , since among these curves we find in particular the curves $L + 2C_1$. We conclude that the system $\Sigma^{(1)}$, just defined, possesses a *total* fundamental curve of the type

$$(8) \quad L^{(1)} = k_2^{(1)}L_2 + k_3^{(1)}L_3 + \cdots + k_s^{(1)}L_s, \quad (s' \leq s),$$

where $L_2 + \cdots + L_s = \bar{L}^{(1)}$ and where $k_2^{(1)}, \cdots, k_s^{(1)}$ are positive integers. In a similar manner the remaining maximal connected components $\bar{L}^{(2)}, \cdots, \bar{L}^{(s)}$ of $L_2 + \cdots + L_s$ give rise to total fundamental curves $L^{(2)}, \cdots, L^{(s)}$ of $\Sigma^{(1)}$, where $L^{(i)}$ is made up of the irreducible components of $\bar{L}^{(i)}$, each counted to a proper positive multiplicity.

We have seen before that the curves C_1 of Σ_1 meet L_1 in one variable point. These curves correspond to the hyperplane sections of \bar{F} on O . The curves C_1 passing through a fixed point of L_1 form a subsystem of Σ_1 of effective degree less than that of Σ_1 , and hence correspond to the sections of \bar{F} by hyperplanes through O and through another point O_1 of \bar{F} , at finite distance or infinitely near O . However, the point O_1 cannot be at a finite distance from O , since L does not carry fundamental points of the birational transformation between F and \bar{F} and therefore to *every* point of the curve L there corresponds an unique point of \bar{F} . This point must coincide with (or be infinitely near) O , since the curve L is carried by the transformation into the point O . It follows that the curves C_1 on a fixed point of L_1 correspond to the hyperplane sections of \bar{F} passing through O and touching there a fixed tangent line of \bar{F} . Since the curves C_1 meet L_1 in only *one* point, which varies as C_1 varies in Σ_1 , it follows that there is a $(1, 1)$ correspondence between the points of L_1 and the tangential directions of \bar{F} at O . The irreducible component L_1 of L thus corresponds to the first order neighborhood of the point O on \bar{F} . Incidentally this shows that L_1 is a *rational curve*.

We can be more precise and show that not only is L_1 a rational curve but that L_1 is free from multiple points, i. e. $p_1 = [L_1] = 0$. The system Σ_1 cuts out on L_1 a g_1^1 . If L_1 had a point of multiplicity > 1 , the subsystem of Σ_1 through this point would then contain L_1 as a fixed part, and this is impossible, since L_1 is not a fundamental curve of Σ_1 . This property extends

similarly to the curves L_i ($i = 2, \dots, s$). In fact, it will be shown in the next section that any curve L_{i+1} ($i = 1, 2, \dots, s-1$) plays the rôle of L_1 for a conveniently defined exceptional curve $L^{(i)}$ of the first kind [formula (16)]. Hence $p_i = 0$ ($i = 1, 2, \dots, s$), and (6) becomes

$$(6') \quad \sum k_i [(L_i^2) + 2] = 1.$$

The above considerations apply to the system $\Sigma^{(1)}$ as well as to Σ_1 . Let $C^{(1)}$ denote a generic curve of $\Sigma^{(1)}$. We have $C^{(1)} \equiv C + C_1$ and hence the following relations, similar to the relations (3), hold:

$$(9) \quad (C^{(1)} \cdot L_1) = 1, \quad (C^{(1)} \cdot L_i) = 0 \quad (i = 2, \dots, s).$$

The curves $C^{(1)}$ meet L_1 in one variable point, and those which pass through a fixed point of L_1 correspond to the sections of \bar{F} by hyperquadrics passing through O and touching there a fixed tangent line of \bar{F} . Let us now consider the curve $L^{(1)}$, given by (8). In view of (7), $L^{(1)}$ meets L_1 in one point, say $P^{(1)}$. Let $O_1^{(1)}$ be the point of \bar{F} , infinitely near O , which corresponds to $P^{(1)}$. The curves $C^{(1)}$ of $\Sigma^{(1)}$, which are constrained to pass through $P^{(1)}$, necessarily contain the whole curve $L^{(1)}$, since $L^{(1)}$ is a fundamental curve of $\Sigma^{(1)}$. The residual system, which we shall denote by $\Sigma_1^{(1)}$, is irreducible, since it does not possess fixed components ($L^{(1)}$ is a total fundamental curve of $\Sigma^{(1)}$) and is not composed of the curves of a pencil (the system $\Sigma_1^{(1)}$ contains partially the system Σ). Moreover $\Sigma_1^{(1)}$ is of effective degree one less than that of $\Sigma^{(1)}$, since the curves of $\Sigma_1^{(1)}$ correspond to the sections of \bar{F} by the hyperquadrics through O and $O_1^{(1)}$. Hence $L^{(1)}$ is itself an exceptional curve of the first kind. In the birational correspondence between F and \bar{F} to the curve $L^{(1)}$ corresponds the point $O_1^{(1)}$ infinitely near O . The exact nature of this correspondence is yet to be investigated.

In a similar manner it can be proved that the other total fundamental curves $L^{(2)}, L^{(3)}, \dots, L^{(s)}$ of $\Sigma^{(1)}$ are exceptional curves of the first kind and that they correspond to points $O_1^{(2)}, O_1^{(3)}, \dots, O_1^{(s)}$ on \bar{F} , infinitely near and in the first neighborhood of O .

The points $O_1^{(1)}, O_1^{(2)}, \dots, O_1^{(s)}$ are distinct. In fact, the point $O_1^{(a)}$ corresponds, in the (1, 1) correspondence between the directions on \bar{F} about O and the points of L_1 , to the intersection $P^{(a)}$ of $\bar{L}^{(a)}$ or, what is the same, of $L^{(a)}$ with L_1 . These intersections $P^{(1)}, P^{(2)}, \dots, P^{(s)}$ are distinct, since the curves $\bar{L}^{(1)}, \bar{L}^{(2)}, \dots, \bar{L}^{(s)}$ have two by two no points in common.

We now prove that (7) can be replaced by the stronger relation:

$$(10) \quad (L^{(1)} \cdot L_1) = 1.$$

In fact, if $C_1^{(1)}$ denotes a generic curve of $\Sigma_1^{(1)}$, so that $C^{(1)} \equiv C_1^{(1)} + L^{(1)}$, then it follows from (9) that $(L_1 \cdot C_1^{(1)}) = (L_1 \cdot C^{(1)}) - (L_1 \cdot L^{(1)}) = 1 - (L_1 \cdot L^{(1)})$. Since $(L_1 \cdot C_1^{(1)}) \geq 0$ and $(L_1 \cdot L^{(1)}) > 0$, (10) follows. We also have incidentally

$$(10') \quad (C_1^{(1)} \cdot L_1) = 0.$$

Similar relations hold for $L^{(2)}, L^{(3)}, \dots, L^{(s)}$.

4. *The successive fundamental points O, O_1, \dots, O_{s-1} of T .* The points $O_1^{(1)}, O_1^{(2)}, \dots, O_1^{(s)}$, defined in the preceding section, are clearly to be considered as fundamental points, infinitely near O , of the birational transformation between F and \bar{F} . The considerations of the preceding section also show that the presence of two or more distinct fundamental points on \bar{F} in the first neighborhood of O corresponds to the case in which the curve $L_2 + \dots + L_s$ is not connected and consists of two or more maximal connected components. If then we wish to consider first the simplest case in which there is only one fundamental point infinitely near and immediately following O , we must assume that $L_2 + \dots + L_s$ is connected. Let us make this assumption. Then the system $\Sigma^{(1)}$ possesses a total fundamental curve $L^{(1)}$ of the type:

$$(11) \quad L^{(1)} = k_2^{(1)}L_2 + k_3^{(1)}L_3 + \dots + k_s^{(1)}L_s,$$

where $k_2^{(1)}, k_3^{(1)}, \dots, k_s^{(1)}$ are positive integers. The fundamental point on \bar{F} , infinitely near O , will be denoted by O_1 . We may apply to the system $\Sigma^{(1)}$ and to the exceptional curve $L^{(1)}$ the results derived in section 2 for Σ and L . Thus, choosing properly our notation, we have the following relations, analogous to the relations (3), (3') and (4):

$$(12) \quad (C_1^{(1)} \cdot L_2) = 1, \quad (C_1^{(1)} \cdot L_i) = 0, \quad (i = 3, \dots, s);$$

$$(12') \quad k_2^{(1)} = 1;$$

$$(13) \quad (L^{(1)} \cdot L_2) = -1; \quad (L^{(1)} \cdot L_i) = 0, \quad (i = 3, \dots, s).$$

The system $\Sigma^{(1)}$ corresponds to the sections of \bar{F} by the hyperquadrics on O , and the system $\Sigma_1^{(1)}$ corresponds to the sections of \bar{F} by the hyperquadrics on O and O_1 . Let us transform birationally \bar{F} into a surface \bar{F}_1 by referring the above hyperquadrics on O to the hyperplanes of a linear space. There arises a birational transformation T_1 between F and \bar{F}_1 in which to the system $\Sigma^{(1)}$ corresponds the system of hyperplane sections of \bar{F}_1 . In the transformation between \bar{F} and \bar{F}_1 , which is locally quadratic at O , the first neighborhood of O is spread out into the points of a straight line; to the point

(direction), O_1 there will correspond a point O'_1 of this line; and it is clear that the system $\Sigma_1^{(1)}$ goes into the system of hyperplane sections of \bar{F}_1 through O'_1 . For the birational transformation T_1 between F and \bar{F}_1 , the system $\Sigma^{(1)}$ and the exceptional curve $L^{(1)}$ play the same rôle as Σ and L played in the transformation T between F and \bar{F} . If $s > 2$, the transformation T_1 will necessarily possess on \bar{F}_1 fundamental points infinitely near O'_1 and hence the transformation T will possess on \bar{F} fundamental points in the second order neighborhood of O . The number of these fundamental points equals the number of maximal connected components of the curve $L_3 + \dots + L_s$. Again, if we wish to consider the simplest case in which T possesses on \bar{F} only one fundamental point O_2 in the second order neighborhood of O , we must assume that $L_3 + \dots + L_s$ is connected.

Considering the system $\Sigma^{(2)}$ which corresponds on F to the sections of \bar{F}_1 by the hyperquadrics on O'_1 (this system contains all reducible curves $C^{(1)} + C_1^{(1)}$) and assuming that $L_3 + \dots + L_s$ is connected, we have that $\Sigma^{(2)}$ possesses a total fundamental curve of the type

$$L^{(2)} = k_3^{(2)}L_3 + \dots + k_s^{(2)}L_s,$$

and that $L^{(2)}$ is an exceptional curve of the first kind. If $\Sigma_1^{(2)}$ denotes the residual system of $\Sigma^{(2)}$ with respect to $L^{(2)}$ and if $C^{(2)}$ and $C_1^{(2)}$ denote generic curves of $\Sigma^{(2)}$ and $\Sigma_1^{(2)}$ respectively, then, with a proper choice of notation, we have the following relations similar to (12), (12') and (13):

$$(14) \quad (C_1^{(2)} \cdot L_3) = 1, \quad (C_1^{(2)} \cdot L_i) = 0 \quad (i = 4, \dots, s);$$

$$(14') \quad k_3^{(2)} = 1;$$

$$(15) \quad (L^{(2)} \cdot L_3) = -1, \quad (L^{(2)} \cdot L_i) = 0 \quad (i = 4, \dots, s).$$

If $s > 3$ the transformation T between F and \bar{F} possesses on \bar{F} fundamental points in the third order neighborhood of O (arising from fundamental points of T_1 on \bar{F}_1 in the second order neighborhood of O'_1) and there is but one such fundamental point if and only if $L_4 + \dots + L_s$ is connected.

The general procedure is now straight-forward and serves to define the fundamental points of T on \bar{F} in the 1st, 2nd, \dots , $(s-1)$ -th order neighborhoods of O . We assume at present that each fundamental point O_i in the i -th neighborhood of O ($i < s-1$) is followed by just one fundamental point O_{i+1} in the neighborhood of order $i+1$. With these assumptions, T possesses on \bar{F} in addition to O other $s-1$ successive fundamental points O_1, \dots, O_{s-1} , infinitely near O_1 , lying on one irreducible algebroid branch of origin O .

We may define by induction the systems $\Sigma^{(i)}$, $\Sigma_1^{(i)}$, the exceptional curves

$L^{(i)}$ ($i = 0, 1, \dots, s-1$) and state their properties as follows. $L^{(i)}$ is a total fundamental curve of $\Sigma^{(i)}$ and we have

$$(16) \quad L^{(i)} = k_{i+1}^{(i)} L_{i+1} + k_{i+2}^{(i)} L_{i+2} + \dots + k_s^{(i)} L_s,$$

where the k 's are positive integers. $\Sigma_1^{(i)}$ is the residual system of $\Sigma^{(i)}$ (of dimension one less than $\Sigma^{(i)}$) with respect to $L^{(i)}$. We have the following relations, similar to the relations (14), (14'), (15):

$$(17) \quad (C_1^{(i)} \cdot L_{i+1}) = 1, \quad (C_1^{(i)} \cdot L_j) = 0 \quad (j = i+2, \dots, s);$$

$$(17') \quad k_{i+1}^{(i)} = 1;$$

$$(18) \quad (L^{(i)} \cdot L_{i+1}) = -1; \quad (L^{(i)} \cdot L_j) = 0 \quad (j = i+2, \dots, s).$$

Here $C^{(i)}$ and $C_1^{(i)}$ denote total curves of $\Sigma^{(i)}$ and $\Sigma_1^{(i)}$ respectively. The curve $L_{i+2} + \dots + L_s$ is connected. The system $\Sigma^{(i+1)}$ is the minimum linear system containing as total curves all reducible curves $C^{(i)} + C_1^{(i)}$.

Let \bar{F}_i be the birational transform of the surface F obtained by referring the curves of $\Sigma^{(i)}$ to the hyperplanes of a linear space, and let T_i denote the birational transformation between F and \bar{F}_i ($\bar{F}_0 = \bar{F}$, $\Sigma^{(0)} = \Sigma$, $T_0 = T$). The system $\Sigma_1^{(i)}$ corresponds by T_i to the system of hyperplane sections of \bar{F}_i on a fixed point $O_i^{(i)}$, while the curves of $\Sigma^{(i+1)}$ correspond to the sections of \bar{F}_i by the hyperquadrics through $O_i^{(i)}$. It is clear that \bar{F}_{i+1} is a birational transform of \bar{F}_i , obtained by referring the sections of \bar{F}_i by the hyperquadrics through $O_i^{(i)}$ to the hyperplane sections of \bar{F}_{i+1} . In this transformation, which is locally quadratic at $O_i^{(i)}$, the point $O_{i+1}^{(i+1)}$, corresponds to a direction about $O_i^{(i)}$. The points $O_0^{(0)} = O$, $O_1^{(1)}$, $O_2^{(2)}$, \dots , $O_{s-1}^{(s-1)}$ are the transforms of the s fundamental points O , O_1 , \dots , O_{s-1} on \bar{F} by the above successive locally quadratic transformations and serve to define the position of the fundamental points O , O_1 , \dots , O_{s-1} on \bar{F} on algebroid branches: every branch on \bar{F} passing through O , O_1 , \dots , O_i corresponds to a branch on \bar{F}_i passing through $O_i^{(i)}$.

The following intersection formulas will be useful in the sequel and are obtained by applying the formulas (10) and (10') to the exceptional curves $L^{(i)}$:

$$(19) \quad (L^{(i)} \cdot L_i) = 1;$$

$$(19') \quad (C_1^{(i)} \cdot L_i) = 0 \quad (i = 1, \dots, s-1).$$

Remark. Our notation implies a definite ordering of the irreducible components L_1, L_2, \dots, L_s of L . From (19) it follows that each curve L_i

intersects one and only one of the curves L_{i+1}, \dots, L_s . If L_{i+a} ($a > 0$) is the curve which L_i intersects, then $(L_i \cdot L_{i+a}) = 1$ and L_{i+a} must be a simple component of $L^{(i)}$. It does not, however, necessarily coincide with the simple component L_{i+1} , as we shall see in the sequel. At any rate, all the intersection numbers $(L_i \cdot L_j)$, $i \neq j$, are either 0 or 1.

5. *The correspondence between L_i and the immediate neighborhood of O_{i-1} .* It has been proved in section 3 that, in the birational correspondence T between F and \bar{F} , the irreducible component L_1 of L corresponds to the first order neighborhood of the fundamental point O on \bar{F} . In a similar manner; in the birational correspondence T_{i-1} between F and \bar{F}_{i-1} the irreducible component L_i of $L^{(i-1)}$ corresponds to the first order neighborhood of the fundamental point $O_{i-1}^{(i-1)}$ on \bar{F}_{i-1} . Hence, to a branch δ_{i-1} on \bar{F}_{i-1} of origin $O_{i-1}^{(i-1)}$ and possessing at $O_{i-1}^{(i-1)}$ a principal tangent line distinct from the singular direction at $O_{i-1}^{(i-1)}$ (that direction to which there corresponds the point $O_i^{(i)}$ on \bar{F}_i), there corresponds a branch γ_{i-1} on F meeting L_i at point P_i . The branch δ_{i-1} arises, by the sequence of locally quadratic transformations applied to $\bar{F}, \bar{F}_1, \dots$, from a branch $\bar{\gamma}_{i-1}$ passing through O, O_1, \dots, O_{i-1} but not through O_i . Hence we have the following

THEOREM 2. *To a branch $\bar{\gamma}_{i-1}$ on \bar{F} , passing through O, O_1, \dots, O_{i-1} but not through O_i , there corresponds a branch γ_{i-1} on F meeting L_i . As we approach on \bar{F} the point O along such a branch $\bar{\gamma}_{i-1}$ the homologous point on F approaches a point P_i on L_i . This point P_i varies as the point on $\bar{\gamma}_{i-1}$ which immediately follows O_{i-1} varies.*

It will be noted that an apparent discontinuity arises as $\bar{\gamma}_{i-1}$ approaches a branch $\bar{\gamma}_i$ passing through O, O_1, \dots, O_{i-1} and O_i . The branch γ_i , which corresponds to $\bar{\gamma}_i$ on \bar{F} , is a branch meeting L_{i+1} but not necessarily L_i . The real situation is the following: as $\bar{\gamma}_{i-1}$ approaches a branch $\bar{\gamma}_i$, the corresponding branch γ_{i-1} on F approaches a locus which degenerates into the curve $L^{(i)}$ and into the branch γ_i .

Another paradoxical circumstance is the following. We have seen that there is a (1, 1) correspondence between the points of L_i and the points on \bar{F} infinitely near and immediately following O_{i-1} . Among these points there is the fundamental point O_i , and it seems natural to expect that the corresponding point on L_i is a common point of L_i and L_{i+1} . However, this is not always the case, as it may well happen that L_i and L_{i+1} do not intersect at all. We shall deal with the general case in the next section. Here an example may suffice. Let $s = 3$ and let the fundamental points O, O_1, O_2 on \bar{F} lie on a

cuspidal branch. The general results of the following section show that in this case we have: $L = L_1 + L_2 + 2L_3$, $(L_1 \cdot L_2) = 0$, $(L_1 \cdot L_3) = 1$, $(L_2 \cdot L_3) = 1$ and incidentally $(L_1^2) = -3$, $(L_2^2) = -2$, $(L_3^3) = -1$. Thus the curve L_1 does not meet L_2 , although there is a point on L_1 which corresponds formally to the direction $O O_1$. This point is the intersection of L_1 and L_3 . If we consider a branch on F passing through this point and if we assume for simplicity that this branch is linear, we find that the corresponding branch on \bar{F} passes *triply* through O and *simply* through O_1 and O_2 .

6. *The geometric significance of the coefficients $k_1^{(j)}$.* A branch $\bar{\gamma}_{i-1}$ on \bar{F} of *smallest order*, passing through O, O_1, \dots, O_{i-1} , arises from a linear branch δ_{i-1} on \bar{F}_{i-1} , of origin $O_{i-1}^{(i-1)}$ and possessing at $O_{i-1}^{(i-1)}$ a generic tangent line. To such a branch there corresponds on F a *linear* branch γ_{i-1} of origin P_{i-1} , a generic point of L_i and therefore not on L_j , $j \neq i$.* It is well known from the theory of singularities that branches $\bar{\gamma}_{i-1}$ of lowest order are also characterized by the fact that such a branch $\bar{\gamma}_{i-1}$ passes *simply* through O_{i-1} and that the points which follow O_{i-1} on $\bar{\gamma}_{i-1}$ are all *free points*. (³, pp. 365-366, 372).

The orders of multiplicity with which $\bar{\gamma}_{i-1}$ passes through the points O, O_1, \dots, O_{i-1} can be obtained as follows. The order of the multiple point O_{j-1} ($j \leq i$) equals the multiplicity with which the branch δ , which corresponds on the surface \bar{F}_{j-1} to $\bar{\gamma}_{i-1}$, passes through the point $O_{j-1}^{(j-1)}$, its origin. We may assume that δ is the branch of some irreducible algebraic curve \bar{D} on \bar{F}_{j-1} and that \bar{D} does not possess other branches of origin $O_{j-1}^{(j-1)}$, distinct from δ . Then we may calculate the multiplicity of \bar{D} (i. e. of δ) at $O_{j-1}^{(j-1)}$ as the difference between the number of intersections of a variable hyperplane section of \bar{F}_{j-1} with \bar{D} and the number of variable intersections with \bar{D} of a hyperplane section through $O_{j-1}^{(j-1)}$. Going back to the surface F we see that the required multiplicity equals $(C^{(j-1)} \cdot D) - (C_1^{(j-1)} \cdot D)$, where D on F is the transform of \bar{D} on \bar{F}_{j-1} . Now $C^{(j-1)} = C_1^{(j-1)} + L^{(j-1)}$ and hence $(C^{(j-1)} \cdot D) - (C_1^{(j-1)} \cdot D) = (L^{(j-1)} \cdot D) = k_i^{(j-1)}$, since the branch γ_{i-1} , and hence also D , has a simple intersection with the curve L_i , does not meet any other irreducible component of $L^{(j-1)}$, and since L_i is a $k_i^{(j-1)}$ -fold component of $L^{(j-1)}$. We thus have the following

THEOREM 3. *The fundamental points O, O_1, \dots, O_{i-1} ($i \leq s$) on \bar{F} lie*

* The truth of this statement follows immediately if we recall that to the sections of \bar{F}_{i-1} by hyperplanes through $O_{i-1}^{(i-1)}$ there correspond on F the curves $C_1^{(i-1)}$ of the system $\Sigma_1^{(i-1)}$ and that by (17) $(C_1^{(i-1)} \cdot L_i) = 1$.

on a branch $\bar{\gamma}_{i-1}$ of lowest order, passing through these points with the multiplicities $k_i, k_i^{(1)}, k_i^{(2)}, \dots, k_i^{(i-1)} = 1$.

As a corollary we have the following inequalities:

$$(20) \quad k_i \geq k_i^{(1)} \geq k_i^{(2)} \geq \dots \geq 1.$$

In particular the branch of lowest order passing through all the s fundamental points O, O_1, \dots, O_{s-1} has at these points the multiplicities $k_s, k_s^{(1)}, k_s^{(2)}, \dots, k_s^{(s-1)} = 1$ respectively.

The branch $\bar{\gamma}_i$ passes through O, O_1, \dots, O_{i-1} and in addition through O_i . The multiplicity of any point $O_j, j \leq i-1$, on this branch cannot be less than its multiplicity for the branch $\bar{\gamma}_{i-1}$ of lowest order passing through O, O_1, \dots, O_{i-1} . Hence we also have the following inequalities:

$$(21) \quad k_j^{(j-1)} \leq k_{j+1}^{(j-1)} \leq k_{j+2}^{(j-1)} \leq \dots \leq k_s^{(j-1)}.$$

7. *The determination of the intersection numbers $(L_i \cdot L_j)$.* In his theory of plane singularities, Enriques (³, p. 381) has introduced the notion of *proximate points* on an algebroid branch. Let O, O_1, \dots, O_{s-1} be a sequence of infinitely near points on an algebroid branch γ , and let ν_i be the multiplicity of O_i on γ . We have for any $i, \nu_i \geq \nu_{i+1}$. If $\nu_i = \nu_{i+1}$, then the set of proximate points of O_i on γ consists by definition of the single point O_{i+1} . If $\nu_i > \nu_{i+1}$ and if $\nu_i = h\nu_{i+1} + \nu'$ ($0 \leq \nu' < \nu_{i+1}$), then it can be proved (³, p. 381) that O_i is followed immediately on γ by h successive ν_{i+1} -fold points $O_{i+1}, O_{i+2}, \dots, O_{i+h}$ and by one ν' -fold point O_{i+h+1} ($i+h+1 \leq s-1$), if $\nu' > 0$. These $h+1$ points constitute the set of proximate points of O_i . If $\nu' = 0$, the set of proximate points of O_i consists of h points O_{i+1}, \dots, O_{i+h} . From the above definition one deduces the following fundamental property of the set of proximate points of a given point: *the multiplicity of O_i on γ equals the sum of the multiplicities of its proximate points.*

We shall have occasion to use the following property of proximate points:

THEOREM 4. *If O_{i+a} is in the set of proximate points of O_i on a given branch γ , then it also belongs to the set of proximate points of O_i on any other branch passing through O_{i+a} (and hence also through O_i).*

This theorem is implicitly contained in Enriques and follows from the construction of the branches of lowest order passing through a given set of successive infinitely near points $O, O_1, \dots, O_i, O_{i+1}, \dots, O_{s-1}$ on a given branch γ . If $O_{i+1}, \dots, O_{i+\beta}$ is the set of proximate points of O_i on γ , it is found that on a branch of lowest order passing through O, O_1, \dots, O_{s-1} the

set of proximate points of O_i consists of the set $O_{i+1}, \dots, O_{i+\beta}$ if $s-1 > i+\beta$, and includes the set O_{i+1}, \dots, O_{s-1} if $s-1 \leq i+\beta$ (*, Chap. I, 2). From this the above theorem follows immediately. We shall now prove the following

THEOREM 5. *Let γ_{s-1} be the branch of lowest order which passes through the infinitely near fundamental points O, O_1, \dots, O_{s-1} on \bar{F} and let the set of proximate points of O_{i-1} on γ_{s-1} ($i < s$) consist of α_i points $O_i, O_{i+1}, \dots, O_{i+\alpha_i}$.* Then $(L_i \cdot L_{i+\alpha_i}) = 1$, $(L_i \cdot L_j) = 0$ if $j > i$ and $j \neq i + \alpha_i$. Moreover the virtual degree (L_i^2) of L_i equals $-(1 + \alpha_i)$.*

For the proof we observe first of all that it is sufficient to prove the theorem for $i=1$ ($O_{i-1} = O, L_i = L_1$), for then we may apply the result to any exceptional curve $L^{(i-1)}$ and to the sequence of fundamental points $O_{i-1}^{(i-1)}, O_i^{(i-1)}, \dots, O_{s-1}^{(i-1)}$ on \bar{F}_{i-1} of the birational correspondence between F and \bar{F}_{i-1} . These points lie on a branch of lowest order, which is the transform of the branch γ_{s-1} , and the proximate points of $O_{i-1}^{(i-1)}$ are clearly the points $O_i^{(i-1)}, \dots, O_{i+\alpha_{i-1}}^{(i-1)}$.

We first prove the theorem when $\alpha = 1$, i. e. when the set of proximate points of O consists of the single point O_1 . By Theorem 3 we have then $k_s = k_s^{(1)}$. Applying Theorem 4, we see that the set of proximate points of O on any branch γ_{i-1} of lowest order passing through O, O_1, \dots, O_{i-1} will consist of the single point O_1 , and hence, again by Theorem 3, we have $k_i^{(1)} = k_i$ ($i = 2, 3, \dots, s$), i. e. $L = L_1 + L^{(1)}$. By (18) we have $(L \cdot L_2) = 0$, $(L^{(1)} \cdot L_2) = -1$, hence $(L_1 \cdot L_2) = 1$ and this proves the first part of the theorem. We also have $(L \cdot L_1) = (L_1^2) + (L^{(1)} \cdot L_1)$ and by (18) and (19) $(L \cdot L_1) = -1$, $(L^{(1)} \cdot L_1) = 1$. Consequently $(L_1^2) = -2$, q. e. d.

By section 4, Remark, L_1 meets only one of the curves L_2, \dots, L_s and the corresponding intersection number is 1. To prove our theorem for α arbitrary, $\alpha > 1$, we shall assume that $(L_1 \cdot L_{\alpha+1}) = 1$ and we shall show that then the set of proximate points of O on the branch γ_{s-1} consists of the points O_1, \dots, O_α and that $(L_1^2) = -(1 + \alpha)$.

We have $(L^{(1)} \cdot L_1) = k_{\alpha+1}^{(1)}$ and hence by (19) $k_{\alpha+1}^{(1)} = 1$. Using the inequalities (20) we find $k_{\alpha+1}^{(2)} = k_{\alpha+1}^{(3)} = \dots = k_{\alpha+1}^{(\alpha)} = 1$, and from this it follows, in view of the inequalities (21), that all the coefficients $k_i^{(j)}$ ($j = 1, 2, \dots, \alpha, i = j+1, \dots, \alpha+1$), are 1. For the sake of clearness we write the simplified expression of the curves $L, L^{(1)}, \dots, L^{(\alpha)}$:

* Here necessarily $i + \alpha_i - 1 \leq s - 1$ because on a branch of lowest order through a given sequence of infinitely near points O, O_1, \dots, O_{s-1} , the set of proximate points of any point of the sequence except the last point O_{s-1} is contained in the given sequence.

$$L = L_1 + L_2 + k_3 L_3 + \dots + k_a L_a + k_{a+1} L_{a+1} + k_{a+2} L_{a+2} + \dots$$
$$\begin{array}{ccccccc} L^{(1)} & = & & L_2 + & L_3 + \dots + & L_a + & L_{a+1} + k_{a+2}^{(1)} L_{a+2} + \dots \\ . & . & . & . & . & . & . & . & . & . & . & . \\ L^{(a-1)} & = & & & & L_a + & L_{a+1} + k_{a+2}^{(a-1)} L_{a+2} + \dots \\ L^{(a)} & = & & & & & L_{a+1} + k_{a+2}^{(a)} L_{a+2} + \dots \end{array}$$

We have replaced k_2 by 1, since k_2 and $k_2^{(1)}$ ($=1$) give the multiplicities of O and O_1 on the branch of lowest order passing through O and O_1 , and hence clearly $k_2=1$. From Theorem 3 it follows that the branch γ_a of lowest order passing through O, O_1, \dots, O_a has at these points the multiplicities $k_{a+1}, 1, \dots, 1$. Consequently on this branch the set of proximate points of O_i , $1 \leq i \leq \alpha-1$, consists of the single point O_{i+1} . It follows by Theorem 4 that also on γ_{s-1} the set of proximate points of O_i , where now $1 \leq i \leq \alpha-2$, consists of the single point O_{i+1} . Since we have already proved our theorem for the case $\alpha=1$, we conclude that the following relations hold:

$$(22) \quad (L_i \cdot L_{i+1}) = 1$$

$$(22') \quad (L_i^2) = -2 \quad (i = 2, 3, \dots, \alpha - 1).$$

By (18) we have $(L \cdot L_1) = -1$ and $(L \cdot L_i) = 0, i > 1$. Using the relations (22), (22') and recalling that by section 4, Remark, each curve L_i intersects only one of the components $L_j, j > i$, we find

$$(23) \quad (L_1^2) \vdash k_{a+1} = -1;$$

$$(23') \quad -2k_2 + k_3 = 0;$$

$$(23'') \quad k_{i-1} - 2k_i + k_{i+1} = 0 \quad (i = 3, \dots, \alpha - 1).$$

Since $k_1 = k_2 = 1$ the relations (23') and (23'') give the following values for k_3, k_4, \dots, k_a :

$$(24) \quad k_3 = 2, k_4 = 3, \dots, k_\alpha = \alpha - 1.$$

The curve L_a intersects one (and only one) of the curves L_{a+1}, L_{a+2}, \dots . Let $(L_a \cdot L_{a+j}) = 1, j \geq 1$. From $(L^{(i)} \cdot L_a) = 0$ ($i = 1, 2, \dots, \alpha - 2$), $(L^{(\alpha-1)} \cdot L_a) = -1$ and $(L^{(\alpha)} \cdot L_a) = 1$ we derive the following relations:

$$(25) \quad 1 + (L_a^2) + k_{a+i}^{(i)} = 0 \quad (i = 1, 2, \dots, \alpha - 1);$$

$$(25') \quad k_{a+i}^{(a)} = 1.$$

From these relations it follows that $k_{a+j}^{(1)} = k_{a+j}^{(2)} = \dots = k_{a+j}^{(a-1)}$ and $k_{a+j}^{(a)} = k_{a+j}^{(a+1)} = \dots = k_{a+j}^{(a+j-1)} = 1$ by (20). Hence the branch γ_{a+j-1} of lowest order passing through the points O, O_1, \dots, O_{a+j-1} has at these points the following multiplicities: $O^{ka+j}, O_1^k, O_2^k, \dots, O_{a-1}^k, O_a^1, \dots, O_{a+j-1}^1$, where we have put $k = k_{a+j}^{(1)}$. We also have $(L \cdot L_a) = 0$ and hence from (24)

$$(\alpha - 2) + (\alpha - 1)(L_a^2) + k_{a+j} = 0.$$

Substituting for (L_a^2) the value $-k-1$ from (25), we find

$$k_{a+j} = (\alpha - 1)k + 1.$$

This value of k_{a+j} shows that on the above branch γ_{a+j-1} the set of proximate points of O consists of the points O_1, O_2, \dots, O_a . By Theorem 4 it follows then that also on the branch γ_{s-1} the points O_1, \dots, O_a constitute the set of proximate points of O , and this proves the first part of the theorem.

We have already observed that the branch γ_a of lowest order passing through the points O, O_1, \dots, O_a possesses at these points the following multiplicities: $O^{k_{a+1}}, O_1^1, \dots, O_a^1$. Also on this branch the points O_1, \dots, O_a must constitute the set of proximate points of O . Hence $k_{a+1} = \alpha$ and by (23), $(L_1^2) = -(1 + \alpha)$, q. e. d.

Remark 1. By Theorem 1, the curve L_s is of virtual degree -1 because $L_s = L^{(s)}$, i. e. L_s is itself an exceptional curve of the first kind.

Remark 2. The preceding theorem shows that the intersection numbers $(L_i \cdot L_j)$ are completely determined by the characters of the branch γ_{s-1} of lowest order passing through the infinitely near fundamental points O, \dots, O_{s-1} . Conversely, it is not difficult to show that *the intersection numbers $(L_i \cdot L_j)$ completely determine the characters of the branch γ_{s-1} , i. e. the multiplicities of the points O, O_1, \dots, O_{s-1} on this branch.* In fact, $-(L_i^2) - 1$ gives the number of proximate points of O_i ($i < s-1$), and hence the multiplicity of O_i on γ_{s-1} can be found if the multiplicities of the points $O_{i+1}, O_{i+2}, \dots, O_{s-1}$ are known, because the multiplicity of O_i equals the sum of the multiplicities of the proximate points. Since the multiplicity of O_{s-1} is 1, the multiplicities of O_{s-2}, O_{s-3}, \dots can be determined step by step.

In a similar manner the multiplicities of the points O, O_1, \dots, O_{i-1} on the branch γ_{i-1} of lowest order containing them can be determined. Hence, by Theorem 3, it follows that *the intersection numbers $(L_i \cdot L_j)$ determine uniquely all the integers $k_i^{(j)}$.*

Remark 3. The determinant $\Delta_i = |(L_\alpha \cdot L_\beta)|$ ($\alpha, \beta = i, i+1, \dots, s$), equals $(-1)^{s-i+1}$. In fact, if we multiply in Δ_1 the second column by k_2 , the third column by k_3, \dots , the s -th column by k_s and add to the first column, we find, in view of equations (5), $\Delta_1 = -\Delta_2$. In a similar manner we find $\Delta_2 = -\Delta_3 = \Delta_4 = \dots = \pm \Delta_s$. Since $\Delta_s = (L_s^2) = -1$, the statement follows. The equations (5) determine the integers k_i in terms of the intersection numbers $(L_i \cdot L_j)$. This is in agreement with the preceding remark.

8. *Free points and satellites.* Our analysis of reducible exceptional curves runs parallel to the analysis of plane singularities due to Enriques. It illustrates very concretely, by means of the intersection properties of the components of the exceptional curve, the notions and conventions used in this theory, as for instance the notion of proximate points. Another concept, also due to Enriques (³, pp. 365-366, 372), is that of *free points and satellites* in a sequence of infinitely near points on an algebroid branch. We shall now show that the knowledge of the intersection numbers $(L_i \cdot L_j)$ allows us to decide whether the fundamental point O_i on γ_{s-1} , which corresponds to a given component L_{i+1} of L , is a free point or a satellite. For this we quote a few properties relative to the classification of the points O, O_1, \dots, O_{s-1} into free points and satellites (⁴, Chap. I, 2).

Let $O_{k+1}, O_{k+2}, \dots, O_{k+l}$ be a sequence of free points such that O_k (if $O_{k+1} \neq O$) and O_{k+l+1} (if $k+l < s-1$) are satellites, and let v_i denote the multiplicity of O_i on γ_{s-1} . Such a sequence enjoys the following characteristic properties:

(α) $v_{k+1} = v_{k+2} = \dots = v_{k+l-1}$; $v_{k+l} < v_{k+l-1}$ except when $k+l = s-1$, in which case $v_{k+l-1} = v_{k+l} = 1$.

(β) If O_{k+1} is distinct from O , then $k \geq 2$ and $v_{k-1} = v_k$. Moreover if $v_k = v_{k-1} = \dots = v_{k-i+1}$ and $v_k \neq v_{k-i}$, then $v_{k-i} = iv_k$.

Recalling the properties of proximate points quoted in section 7, we conclude from (α) and (β) that

(1) Every satellite is the last proximate point of at least one point preceding it.

(2) If O_k is a satellite followed by a free point O_{k+1} , then O_k is the last proximate point of two points O_{k-1} and O_{k-2} preceding it, and conversely.

(3) The last proximate point of any point O_i is never a free point followed by a satellite.

From (1), (2) and (3) there follows

THEOREM 6. *If L_{i+1} does not meet any curve L_j , $j < i+1$, then O_i is a free point followed by a satellite O_{i+1} , and conversely. If L_{i+1} meets two of the curves L_j , $j < i+1$, then O_i is a satellite followed by a free point O_{i+1} , and conversely.*

This theorem characterizes the *last* point of a sequence of satellites followed by free points and the *last* point of a sequence of free points followed by satellites and hence enables us to find out the division of the set O, O_1, \dots, O_{s-1} into free points and satellites from the intersection properties of the curves L_1, L_2, \dots, L_s .

We add a few remarks concerning the $(1, 1)$ correspondence between the points of L_{i+1} and the points of the immediate neighborhood of O_i (points in the neighborhood of order $i + 1$ of O). It is not difficult to show (⁴, Chap. I, 2) that the simple infinity of points immediately following O_i contains *one* satellite, if O_i is a free point, and *two* satellites if O_i is itself a satellite. What point or points of L_{i+1} correspond to this satellite or these satellites? To answer this question it is necessary to observe that if \bar{O}_{i+1} is a point in the immediate neighborhood of O_i , then the branch of lowest order passing through $O, O_1, \dots, O_i, \bar{O}_{i+1}$ is of the same order as the branch γ_i of lowest order passing through O, O_1, \dots, O_i if \bar{O}_{i+1} is a free point, and is of higher order than γ_i if \bar{O}_{i+1} is a satellite. A branch γ_i corresponds to a *linear* branch on F , meeting L_{i+1} in a generic point. We obtain on F a branch of higher order than γ_i if and only if that linear branch intersects L_{i+1} in a point where L_{i+1} intersects some further component of L . Recalling that the fundamental point O_{i+1} corresponds formally to the intersection of L_{i+1} with a curve L_j , $j > i + 1$ (section 5), we conclude as follows: if O_i is a free point and O_{i+1} is a satellite, then this satellite corresponds to the intersection of L_{i+1} with an L_j , $j > i + 1$ (it should be noticed that in this case, by Theorem 6, L_{i+1} does not meet any curve L_j , $j < i + 1$). If O_i and O_{i+1} are both free points, then, by Theorem 6, L_{i+1} necessarily intersects one and only one curve L_j , $j < i + 1$; to this point of intersection corresponds *the* satellite which immediately follows O_i . If O_i is a satellite and O_{i+1} is a free point, then, by Theorem 6, L_{i+1} meets two curves L_j , $j < i + 1$, and the two points of intersection correspond to the two satellites in the immediate neighborhood of O_i . Finally if O_i and O_{i+1} are both satellites, then by Theorem 6 L_{i+1} meets only one curve L_j , $j < i + 1$, and the point of intersection corresponds to the second satellite, distinct from O_{i+1} , which immediately follows O_i .

9. *Characterization of reducible exceptional curves of the first kind.* It is known (¹) that the conditions which characterize an *irreducible* exceptional curve L_1 of the first kind, on a surface free from singularities, are the following: $(L_1^2) = -1$, $[L_1] = 0$, these characters being evaluated on the assumption that L_1 is virtually free from base points. The curve L_1 is rational and *free from singularities*, because, if L_1 had singularities, the virtual genus $[L_1]$ would be positive.

We show that the properties of the intersection numbers $(L_i \cdot L_j)$, derived in the preceding section, together with the relations $[L_i] = 0$, proved in section 3, characterize these reducible exceptional curves of the first kind, which arise from birational transformations possessing a sequence of infinitely

near fundamental points lying on an irreducible algebroid branch. In exact terms we prove the following

THEOREM 7. *Let L_1, L_2, \dots, L_s be a set of irreducible curves, virtually free from base points and of virtual genus zero, on a surface F free from singularities, and let there exist an arrangement of these curves, for instance the one written above, such that the following conditions are satisfied: (i) each curve L_i meets one and only one of the curves L_{i+1}, \dots, L_s which follow L_i ; (ii) if L_i meets $L_{i+\alpha}$ ($i < s$), then $(L_i \cdot L_{i+\alpha}) = 1$, $(L_i^2) = -(1 + \alpha)$ and $(L_{i+j} \cdot L_{i+j+1}) = 1$ ($j = 1, 2, \dots, \alpha - 2$); (iii) $(L_s^2) = -1$. Under these conditions a convenient combination $k_1 L_1 + k_2 L_2 + \dots + k_s L_s$ ($k_1 = k_2 = 1$) with positive integral coefficients is an exceptional curve of the first kind. The assumed arrangement L_1, \dots, L_s is uniquely determined and so also are the integers k_i .*

Proof. Since the theorem is true for $s = 1$ ⁽¹⁾, we shall prove it by induction, assuming it to be true for $s - 1$. Since, by hypothesis, $[L_s] = 0$ and $(L_s^2) = -1$, L_s is an exceptional curve of the first kind. There exists therefore a birational transformation of F into a surface \bar{F} , in which to L_s there corresponds a simple point O of \bar{F} and which does not possess fundamental points on L_s . It can also be assumed that the transformation possesses no fundamental points on F , that L_s is the only fundamental curve of the transformation on F and that \bar{F} is free from singularities.* Let $\bar{L}_1, \bar{L}_2, \dots, \bar{L}_{s-1}$ be the transforms of L_1, \dots, L_{s-1} on F . The virtual degree of a curve is invariant under birational transformations. However, if a given L_i meets L_s , then L_i passes through the point O simply, since $(L_i \cdot L_s) = 1$, and this point must be considered as an assigned base point of \bar{L}_i ^(4, Chap. II, 7). If we understand by (\bar{L}_i^2) the virtual degree of \bar{L}_i virtually free from base points, then $(\bar{L}_i^2) = (L_i^2)$ if $(L_i \cdot L_s) = 0$ and $(\bar{L}_i^2) = (L_i^2) + 1$ if $(L_i \cdot L_s) = 1$.

With these preliminaries we proceed to prove that the ordered sequence of curves $\bar{L}_1, \bar{L}_2, \dots, \bar{L}_{s-1}$ on \bar{F} satisfies the conditions for an exceptional curve stated in the above theorem. From (i) we deduce that $(L_{s-1} \cdot L_s) = 1$ and by (ii) $(L_{s-1}^2) = -2$. Hence $(\bar{L}_{s-1}^2) = (L_{s-1}^2) + 1 = -1$ and this shows that the condition (iii) holds. To show that the conditions (i) and (ii) are satisfied, we make the following self-evident remarks:

* The defining linear system Σ on F associated with the transformation, i. e. the system which goes into the system of hyperplane sections of \bar{F} , may be assumed to be free from base points and non-singular ^(14, Chap. IV, 4). If the transformation possesses on F fundamental curves other than L_s and hence fundamental points on \bar{F} distinct and necessarily at a finite distance from O , these fundamental points can be transformed back into curves by means of a sufficiently general linear system on \bar{F} , free from fundamental curves and free from base points outside these fundamental points.

(α) If $(L_i \cdot L_s)$ and $(L_j \cdot L_s)$ are not both 1 ($i, j < s$), then $(\bar{L}_i \cdot \bar{L}_j) = (L_i \cdot L_j)$.

(β) If $(L_i \cdot L_s) = (L_j \cdot L_s) = 1$, then $(\bar{L}_i \cdot \bar{L}_j) = (L_i \cdot L_j) + 1$.

We consider any curve L_i , $i < s-1$. Let $(L_i \cdot L_{i+\alpha}) = 1$, $(L_i \cdot L_j) = 0$, $j > i$ and $j \neq i + \alpha$. If $i + \alpha \neq s$, then by (α) $(\bar{L}_i \cdot \bar{L}_{i+\alpha}) = 1$ and $(\bar{L}_i \cdot \bar{L}_j) = 0$ if $j > i$ and $j \neq i + \alpha$. Moreover by (α) applied to the case $i = j$, it follows that $(\bar{L}_i^2) = (L_i^2) = -1 - \alpha$. If $i + \alpha = s$, then by (β) (where now $j = s-1$) we have $(\bar{L}_i \cdot \bar{L}_{s-1}) = 1$, since $(L_i \cdot L_{s-1}) = 0$. Moreover by (ii) we have $(L_j \cdot L_s) = 0$ ($j = i+1, \dots, s-2$), and hence by (α) $(\bar{L}_i \cdot \bar{L}_j) = 0$ ($j = i+1, \dots, s-2$), since $(L_i \cdot L_j) = 0$. We also have by (β), when we put $i = j$, $(\bar{L}_i^2) = (L_i^2) + 1 = -(s-i) - 1 + 1 = -(s-i-1) - 1$. We have thus proved that every curve \bar{L}_i ($i = 1, 2, \dots, s-2$), meets one and only one of the curves \bar{L}_j , $j > i$, and that if $(\bar{L}_i \cdot \bar{L}_{i+\alpha}) = 1$, $\alpha > 0$, then $(\bar{L}_i^2) = -\alpha - 1$. Finally if $(L_i \cdot L_{i+\alpha}) = 1$, then by (ii) $(L_j \cdot L_{j+1}) = 1$ ($j = i+1, \dots, i+\alpha-2$), and consequently $(\bar{L}_j \cdot \bar{L}_{j+1}) = 1$ for the same values of j . Hence all the conditions (i), (ii), (iii) are satisfied by the ordered sequence $\bar{L}_1, \dots, \bar{L}_{s-1}$.

From the hypothesis that our theorem is true for $s-1$, it follows that a proper combination $\bar{L} = \bar{k}_1 \bar{L}_1 + \dots + \bar{k}_{s-1} \bar{L}_{s-1}$ of the curves \bar{L}_i is an exceptional curve of the first kind. There exists then a linear system $\bar{\Sigma}$ on \bar{F} possessing \bar{L} as a total fundamental curve, satisfying the conditions (i), (ii), and (iii) of the definition given in section 1 and having no base points on \bar{L} . It is seen immediately that the curves L_1, \dots, L_s are fundamental curves of the system Σ on F which corresponds to $\bar{\Sigma}$, that a proper combination $k_1 L_1 + \dots + k_s L_s$ of these curves is a total fundamental curve of Σ (since by the conditions (i), (ii), (iii) of the preceding theorem the curve $L_1 + \dots + L_s$ is connected) and that Σ satisfies the conditions (i), (ii), (iii) of the definition given in section 1. It has been proved already (section 7, Remark 2) that the integers k_i are uniquely determined. It remains to prove that the arrangement L_1, L_2, \dots, L_s is uniquely determined. Let us assume that $L_{i_1}, L_{i_2}, \dots, L_{i_s}$ is another arrangement satisfying the conditions (i), (ii) and (iii), where i_1, i_2, \dots, i_s is a permutation of the indices $1, 2, \dots, s$. Since by (iii) $(L_{i_s}^2) = -1$ and since by (ii) L_s is the only curve whose virtual degree is -1 , it follows that L_{i_s} coincides with L_s . We shall therefore assume that $i_j = j$ for $j = \alpha, \alpha+1, \dots, s$ ($1 < \alpha \leq s$), and we shall prove that $i_{\alpha-1} = \alpha-1$. Let $i_\beta = \alpha-1$, and let $(L_{\alpha-1} \cdot L_{\alpha+\gamma}) = 1$, $\gamma \geq 0$, by (ii). We have, by (ii), $(L_{\alpha-1}^2) = -(2+\gamma)$, and by the same condition (ii), applied to the arrangement $L_{i_1}, L_{i_2}, \dots, L_{i_{\alpha-1}}, L_\alpha, L_{\alpha+1}, \dots, L_s$, we have $(L_{\alpha-1}^2) = (L_{i_\beta}^2) = -(\alpha+\gamma-\beta+1)$. Hence $2+\gamma = \alpha+\gamma-\beta+1$, or $\beta = \alpha-1$, q. e. d.

10. *Fundamental points on several irreducible branches.* In the previous sections we have treated the case of a birational transformation with a succession of infinitely near fundamental points on a single irreducible algebroid branch. We found it necessary to assume that the curves $L_i + \cdots + L_s$ ($i=1, \dots, s$) were connected. When these curves were not connected we saw in section 3 that the aggregate of fundamental points would not lie all on one branch. It is the purpose of this section to consider this latter case.

The curve $L_1 + \cdots + L_s$ is connected (see 2), but we shall suppose that $L_2 + \cdots + L_s$ breaks up into ρ (≥ 1) maximal connected components $\bar{L}^{(1)}, \dots, \bar{L}^{(\rho)}$, having two by two no points in common (see 3). Then on the surface \bar{F} , in the first neighborhood of the point O , we have the ρ distinct fundamental points $O_{1,1}, \dots, O_{\rho,1}$ of the transformation T^{-1} . We shall fix our attention on one of these, say $O_{1,1}$. In the immediate neighborhood of $O_{1,1}$ T^{-1} may have several fundamental points, among which there may or may not be the * proximate point of O following $O_{1,1}$. We shall suppose that T^{-1} has a fundamental point $O_{1,2}$, immediately following $O_{1,1}$, which is a proximate point of O . Then in the immediate neighborhood of $O_{1,2}$, T^{-1} may have several fundamental points, among which there may or may not be the proximate point of O following $O_{1,2}$. We shall suppose that T^{-1} has a fundamental point $O_{1,3}$ immediately following $O_{1,2}$, which is a proximate point of O . We continue in this way to define the sequence of points $O, O_{1,1}, \dots, O_{1,\sigma}$ such that each point is fundamental for T^{-1} and that $O_{1,i}$ is the proximate point of O immediately following $O_{1,i-1}$ ($i=2, \dots, \sigma$).

We shall suppose that the fundamental points of T^{-1} which follow $O_{1,\sigma}$ include no proximate point of O . From the preceding sections it follows that the immediate neighborhood of each point of the sequence $O, O_{1,1}, \dots, O_{1,\sigma}$ is represented in (1, 1) fashion by the points of irreducible components of L , namely $L_1, L_{1,1}, \dots, L_{1,\sigma}$ respectively. Consider now a linear system of curves Φ on \bar{F} , cut out by a system of forms of a sufficiently high order, having base points only at the points $O, O_{1,1}, \dots, O_{1,\sigma}$ and free from fundamental curves on \bar{F} . By relating the system Φ to the hyperplanes of a linear space, we induce a transformation T' and transform \bar{F} into a surface F' ; from the considerations in sections 1-9, it is clear that F' has an exceptional curve of the type treated in those sections. F' is the transform of \bar{F} by the product

* We say "the proximate point of O ," because it can be easily shown that there is only one such point in the immediate neighborhood of $O_{1,1}$. More generally, it is not difficult to prove that if $O_i, O_{i+1}, \dots, O_{i+\beta}$ ($\beta > 0$) is a sequence of successive points in the neighborhoods of order $i, i+1, \dots, i+\beta$ of a point O , such that $O_{i+\beta}$ is a proximate point of O_i , then there is one and only one point $O_{i+\beta+1}$ immediately following $O_{i+\beta}$, which is a proximate point of O_i .

$T \cdot T'$, and this transformation is free from fundamental points on F . The exceptional curve on F' consists of the curves $L'_1, L'_{1,1}, \dots, L'_{1,\sigma}$, the transforms of $L_1, \dots, L_{1,\sigma}$ on F ; the remaining components of L on F are replaced by points of F' .

Now in the exceptional curve on F' , L'_1 must intersect $L'_{1,\sigma}$ by Theorem 5, since on \bar{F} $O_{1,\sigma}$ is the last proximate point of O . Does this imply that the corresponding curves L_1 and $L_{1,\sigma}$ on F intersect? Let us assume that the intersection of L'_1 and $L'_{1,\sigma}$ arises from the intersection of L_1 and $L_{1,\sigma}$ on F with a curve L_β which has been replaced by a point on F' and let O_β be the fundamental point on \bar{F} whose immediate neighborhood is in $(1, 1)$ correspondence with the points of L_β . This point is not in the sequence $O, O_{1,1}, \dots, O_{1,\sigma}$, since O_β is transformed by T' into a point of F' . On the other hand the fact that $L_{1,\sigma}$ and L_β intersect implies that if $L^{(j)}$ ($j = 0, 1, \dots, \sigma - 1$) denotes the exceptional curve on F which corresponds to the point $O_{1,j}$ ($O_{1,j} \equiv O$, $L^{(0)} \equiv L$), then $L_{1,\sigma}$ and L_β belong to one and the same maximal connected component of $L^{(j)} - L_{1,j}$, i. e. to $L^{(j+1)}$. Hence there exists a branch passing through $O, \dots, O_{1,\sigma}; O_\gamma, \dots, O_\beta$. Set up a transformation T'' , as T' was determined, possessing only O, \dots, O_β as fundamental points and transform \bar{F} into F'' ; F'' has an exceptional curve of the type considered in sections 4-9. In this exceptional curve denote by L_j'' the transform of L_j on F . Then $(L''_1 \cdot L''_{1,\sigma}) = 1$ by Theorem 5 and $(L''_1 \cdot L''_\beta) = 1$ since $(L_1 \cdot L_\beta) = 1$ on F and since the transformation TT'' is free from fundamental points on F . But this is impossible, since L''_1 cannot meet two distinct irreducible components of the exceptional curve on F'' .

Hence we must have on F , $(L_1 \cdot L_{1,\sigma}) = 1$. This constitutes an extension of Theorem 5 in whatever way the maximal connected components of $L_i + \dots + L_s$ ($i = 2, \dots, s$) may decompose. In all cases we have shown that L_1 intersects that curve which is the map of the immediate neighborhood of the fundamental point, which is the last proximate point of O .

In Theorem 5 we were also able to derive the virtual degree of L_1 and hence the virtual degree of any component of L ; we found $(\bar{L}_1^2) = -(1 + \text{number of proximate points of } O)$; we wish to prove this for the present case when the fundamental points lie on several irreducible branches.*

The curve $\bar{L} = L_1 + \dots + L_s$ is connected, but $L_2 + \dots + L_s$ breaks up into $\rho \geq 1$ maximal connected components $\bar{L}^{(1)}, \dots, \bar{L}^{(\rho)}$; L_1 meets $\bar{L}^{(\beta)}$ ($\beta = 1, \dots, \rho$) in one point each, and each $\bar{L}^{(\beta)}$ gives rise to an exceptional curve $\bar{L}^{(\beta)}$ and to a fundamental point $O_{\beta,1}$ in the first neighborhood of O .

* The set of proximate points of a given point O , when several branches pass through O , is by definition the set of all the proximate points of O on the different branches through O (4, Chap. I, 2).

We shall fix our attention on one of these points, say $O_{1,1}$, and on the corresponding curve $\bar{L}^{(1)} = L_2 + \cdots + L_s$. The intersection $(L_1 \cdot \bar{L}^{(1)}) = 1$ arises from the intersection of L_1 with some curve L_{a_a} of $\bar{L}^{(1)}$ such that $(L_1 \cdot L_{a_a}) = 1$. Now $\bar{L}^{(1)} - L_2$ decomposes into several maximal connected components, one of which contains L_{a_a} ; let this component be $\bar{L}_{a_2}^{(2)} = L_{a_2} + \cdots + L_{a_a} + \cdots$; we have $(L_2 \cdot \bar{L}_{a_2}^{(2)}) = 1$ [as $(L_1 \cdot \bar{L}^{(1)}) = 1$] and $\bar{L}_{a_2}^{(2)}$ gives rise to a fundamental point $O_{1,2}$ in the immediate neighborhood of $O_{1,1}$. Then $\bar{L}_{a_2}^{(2)} - L_{a_2}$ breaks up into several maximal connected components, one of which contains L_{a_a} ; let this component be $\bar{L}_{a_3}^{(3)} = L_{a_3} + \cdots + L_{a_a} + \cdots$. As before, there is a fundamental point $O_{1,3}$ in the immediate neighborhood of $O_{1,2}$. We continue in this manner and come finally to a fundamental curve $\bar{L}_{a_a}^{(a)} = L_{a_a} + \cdots$ and this gives rise to a fundamental point $O_{1,a}$ in the neighborhood of order α of O . From the proof in the earlier part of this section it follows that $O_{1,a}$ is the last proximate point of O after $O_{1,1}$ among the fundamental points of T^{-1} .

The proof of section 5 extends directly to show that the immediate neighborhood of $O_{1,a}$ is in $(1, 1)$ correspondence with the points of the curve L_{a_a} ; then the proof of section 6 follows and shows that the branch of minimum order through $O, O_{1,1}, \cdots, O_{1,a}$ has at these points the multiplicities $k_{a_a}, k_{a_a}^{(1)}, \cdots, k_{a_a}^{(a)} = 1$, where these values are the coefficients of L_{a_a} in the exceptional curves $L, L^{(1)}, L_{a_2}^{(2)}, \cdots, L_{a_a}^{(a)}$, derived from the fundamental curves $\bar{L}, \bar{L}^{(1)}, \cdots, \bar{L}_{a_a}^{(a)}$. But $k_{a_a}^{(1)} = 1$ by (10) of section 3. Therefore $k_{a_a}^{(2)} = \cdots = k_{a_a}^{(a)} = 1$. Since the multiplicity of a point is equal to the sum of the multiplicities of its proximate points, we have at once $k_{a_a} = \alpha$. We apply this procedure to the curves $L^{(\beta)}$ ($\beta = 1, \cdots, \rho$), and from the relation $(L \cdot L_1) = -1$ we derive in all cases the desired result: $(L_1^2) = -(1 + \text{number of proximate points of } O)$. As in Theorem 5, the proof extends to the other irreducible components of L .

REFERENCES.

- ¹ G. Castelnuovo and F. Enriques, "Sopra alcune questioni fondamentali nella teoria delle superficie algebriche," *Annali di Matematica pura ed applicata*, Ser. 3, vol. 6 (1901), pp. 165-225.
- ² F. Enriques, "Sulla proprietà caratteristica delle superficie algebriche irregolari," *Rendiconti della Reale Accademia delle scienze di Bologna*, nuova serie, vol. 9 (1904-05).
- ³ F. Enriques and O. Chisini, *Lezioni sulla teoria geometrica delle equazioni e delle funzioni algebriche*, vol. 2.
- ⁴ O. Zariski, *Algebraic Surfaces*, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. III, 5, Berlin, Springer (1934).

THE TOPOLOGICAL TRANSFORMATIONS OF A SIMPLE CLOSED CURVE INTO ITSELF.

By E. R. VAN KAMPEN.

I. The theory of a topological transformation of a simple closed curve into itself has been developed first of all by Poincaré (a), who needed it for his investigations on the differential equations on a torus. Later P. Bohl (b) formulated a result closely related to the sufficiency in our (13). Then H. Kneser (c) and J. Nielsen (d) gave improved proofs of the existence of the number τ (3) and of the transformation functions $y = g(x)$ and $x = h(y)$ (V). In this paper we have used Nielsen's proof that was published in the Danish language. Denjoy (e, f) found his sufficient condition for the transitive case (15) using continued fractions. Kneser (g) published a theorem equivalent with the necessity in (13). Wintner and Lewis (h) laid the foundation for (14). The papers a, b, f, g consider the question as a problem for differential equations on the torus.

In this paper we give the theory for a simple closed curve only, considerably simplified and completed on several points. The consequences for the solutions of a differential equation on the torus are immediate. The examples that have been assembled will not be repeated here. An account will be found in Denjoy's paper.

II. We consider a simple closed curve C and on it a coördinate x taking all real values, such that two values of x correspond to the same point of C if and only if their difference is an integer.

If a topological transformation of C into itself changes the orientation of C then the transformation has exactly two fixpoints and interchanges the two arcs determined by the fixpoints on C . We exclude this case completely.

Any other topological transformation T of C into itself can be determined by a monotonically increasing continuous real function $f(x)$, satisfying the relation $f(x+1) = f(x) + 1$. However if T is known the function $f(x)$ is not uniquely determined but can be replaced by $f(x) + n$, where n is an arbitrary integer. We can consider $f(x)$ as a topological transformation of the line of real numbers into itself.

If we denote by T^n and f^n the n -th powers of the transformations T and f , then we can represent T^n by any of the functions $f^n(x) + k$, where k is an arbitrary integer.

If p is any point of C and x any real number we will sometimes write p_n for $T^n(p)$ and x_n for $f^n(x)$. As T^n is again a topological transformation of C into itself, $f^n(x)$ will be the same type of real function as $f(x)$.

III. In this section we introduce a real number τ characterizing in a very definite topologically invariant sense the average size of the transformation affected by T on the points of C .

(1) *The limit $\lim_{|n| \rightarrow \infty} (1/n)x_n = \tau$ exists and is independent of x .*

As soon as we know that the limit exists for a particular value of x , then the rest of the theorem is very simple. For

$$|x_n - y_n| < |x - y| + 1, \text{ so } \lim_{|n| \rightarrow \infty} (1/n) \cdot (x_n - y_n) = 0.$$

If some power T^m of T has a fixpoint with the coördinate y , then we can find an integer p such that $y_m = y + p$. Using the formula $f^m(x+1) = f^m(x) + 1$ we can prove by induction $y_{mn} = y + pn$ and a simple limiting process shows that in this case τ exists and is equal to the rational number p/m .

If no power of T has a fixpoint then *no integers m and p and real number x exist such that $x_m = x + p$* . Thus we can find corresponding to any integer m an integer p such that $p + x < x_m < p + x + 1$, for all real x . In fact we only have to find the value of p for one particular value of x and then the inequalities are true for all x because the corresponding equalities never take place. Applying the inequality for $x = 0, 0_m, 0_{2m}, \dots, 0_{m(n-1)}$ and adding we find:

$$np < 0_{mn} < n(p + 1).$$

Comparing this with the original inequality for $x = 0$ we can prove:

$$\left| \frac{1}{mn} 0_{mn} - \frac{1}{n} 0_n \right| < \frac{1}{|m|}.$$

It is clear how this reasoning has to be rearranged if n is negative. If we interchange m and n in the last formula we can conclude:

$$\left| \frac{1}{n} 0_n - \frac{1}{m} 0_m \right| < \frac{1}{|m|} + \frac{1}{|n|},$$

from which the existence of τ follows immediately.

(2) *τ is irrational if and only if no power of T has a fixpoint.*

If τ is the invariant of $f(x)$, then $\tau m + k$ is the invariant of $f^m(x) + k$.

So if τ is rational we can find a power $S = T^m$ of T and a corresponding function $g(x) = f^m(x) + k$ such that $\lim_{|n| \rightarrow \infty} (1/n)g^n(0) = m\tau + k = 0$.

Supposing at the same time that S has no fixpoints, we can assume for reasons of symmetry that $g(x) > x$ for all x . This makes $g_n(0)$ an increasing function of n . But we must have $g_n(0) < 1$, otherwise we could conclude $m\tau + k > 1/n$. So the sequence $g_n(0)$ has a cluster point x and x must necessarily determine a fixpoint of S contrary to our supposition. So τ must be irrational if no power of T has a fixpoint. From now on we will suppose that τ is irrational.

(3) *The number systems $n\tau + m$ and $x_n + m$ are in a one-to-one monotonous correspondence.*

Here x is arbitrary fixed, m and n take all integer values. We note first that the particular value of x chosen has no influence on the order of the elements in $x_n + m$. This depends on the fact that $x_n - x_k$ can never take an integer value. Choosing $x = 0$ we have to prove that $0_n + m < 0_k + l$ and $n\tau + m < k\tau + l$ are equivalent inequalities. The first can be brought into the form $0_{n-k} < l - m$. As in the proof of (1) this leads to $\tau(n - k) < l - m$. And the conclusion is reversible because $0_{n-k} > l - m$ would lead to $\tau(n - k) > l - m$.

(4) *The residual class of τ modulo one is a topological invariant of T and the oriented curve C .*

We know that the systems of numbers $n\tau + m$ and $0_n + m$ are ordered the same way for each coördinate system of C . If we normalize the functions of x belonging to different coördinate systems with the same orientation by the condition $0 < 0_1 < 1$ then the order of the system $0_n + m$ does not depend on the coördinate system chosen. But then the same is the case for the system $n\tau + m$, so the residual class of τ modulo one does not depend on the coördinate system.

IV. We now consider the set $P = \{p_n\}$ formed by all transforms p_n of p and the set E of all cluster points of P .

(5) *E is invariant under the transformation T and independent of p .*

The first part is obvious because P is invariant under T . The second part follows very easily from the following statement:

(6) *The two closed segments e and e^1 determined on C by two points p_m and*

p_n of P both contain at least one of the transforms q_i of any point q of C . To prove this for e we consider the finite sums $e + T^{m-n}e + T^{2(m-n)}e + \dots + T^{k(m-n)}e$. As τ is irrational, so different from zero, these sums of adjacent intervals must fill C completely if k is sufficiently large. So one of the intervals $T^{i(m-n)}e$ must contain q and e must contain $q_{i(n-m)}$.

If now the sequence p_{α_ν} converges to r , then there must be a sequence of transforms q_{b_ν} of q converging to the same point r , so E is independent of p .

(7) E is perfect and either equal to C (transitive case) or nowhere dense on C (intransitive case).

As E is closed the first part follows from (6) if we take q to be a point of E . The second part follows from the fact that P is on C either everywhere dense or nowhere dense. If P is dense in an arbitrary interval of C then we can assume that interval to be e of (6). As a finite number of transforms of that interval fills C completely P must be everywhere dense on C .

V. We now make a closer study of the sets of numbers $A = \{n\tau + m\}$ and $B = \{x'_n + m\}$ mentioned in (3). We take for x' in $x'_n + m$ a fixed but arbitrary number corresponding to one of the points of E .

We are going to define a real function $y = g(x)$. For any number in B we put the function equal to the corresponding number in A . Because of (3) and because A is dense on the real axis the function is continuous and increasing on B . If B is also dense on the real axis (transitive case) this function on B can be extended in a unique way to a continuous increasing function $y = g(x)$, defined for all x . If B is not dense on the real axis (intransitive case) then the function on B can be extended in a unique way to a continuous non-decreasing function $y = g(x)$, constant only on the closed intervals corresponding to the segments of $C - E$.

In the transitive case the function $y = g(x)$ has a continuous increasing inverse $x = h(y)$ transforming the y -axis into the x -axis. In the intransitive case we can define the inverse $x = h(y)$ in such a way that it is increasing and has a countable number of discontinuities corresponding to the countable number of intervals on which $g(x)$ is constant.

The function $y = g(x)$ must satisfy the relations

$$(8) \quad g(x+1) = g(x) + 1,$$

$$(9) \quad gf(x) = g(x) + \tau,$$

for if x is any number in B and y the corresponding number in A then $y + 1$ corresponds to $x + 1$ and $y + \tau$ corresponds to $f(x) = x_1$.

VI. We can choose the variable y , occurring in $y = g(x)$, as a coördinate proportional to the length of arc on some circle D in a euclidean plane. Then (8) shows that $y = g(x)$ defines a univalued and continuous transformation S of C into D , such that S transforms E into D and eventual segments of $C - E$ into points of D . If no such segments exist (transitive case) S must be T a topological transformation.

Now (8) and (9) show that the transformation T of C is transformed by S into a rotation of D over the angle $2\pi\tau$, so finally clearing up completely the meaning of the number τ . If the transformation T is transitive T is topologically equivalent with a rotation of a circle over an angle $2\pi\tau$. If T is intransitive the same equivalence is true after an invariant countable set of closed segments on C has been transformed into a countable set of points by a preliminary transformation of C .

VII. The problem now arises to find conditions for a transformation T of C (of which no power has a fixpoint) to be transitive or intransitive. The N. S. condition for T to be transitive is that the transforms p_n of a point p of C are everywhere dense on C (7). The following condition on the same points is more suitable for applications:

(10) *N. and S. for T to be transitive is for the transforms p_n of a fixed point p to have the following property: Whenever the sequence p_{a_ν} , a_ν integers, $\nu = 1, 2, \dots$, is convergent then the sequences $p_{a_\nu+n}$ depending on n converge uniformly in n .*

The necessity is very simple: If T is transitive the curve C can be obtained as the result of the (uniformly) continuous transformation S^{-1} of D into C (see VI). But if a sequence of numbers τa_ν modulo one converges then $\tau(a_\nu + n)$ modulo one converges uniformly in n , and this uniform convergence is not disturbed by the transformation S^{-1} .

We now prove the sufficiency. If p_{a_ν} converges to an arbitrary point q of E (5), then $\lim_{\nu \rightarrow \infty} p_{a_\nu + b_{\mu+n}} = q_{b_{\mu+n}}$. If we select the sequence b_μ , such that q_{b_μ} converges then $\lim_{\mu \rightarrow \infty} \lim_{\nu \rightarrow \infty} p_{a_\nu + b_{\mu+n}}$ exists. As a consequence of the condition in (10) $p_{a_\nu + b_{\mu+n}}$ will converge uniformly in n and b_μ so $q_{b_{\mu+n}}$ will converge uniformly in n . In other words the condition in (10) is satisfied for any point of E if it is satisfied for p . We apply it to the endpoints q and r of an eventual complementary segment e of E in C . If q_{a_ν} converges then r_{a_ν} must converge to the same point (at least if all a_ν are distinct). But q_{n+a_ν} and r_{n+a_ν} could not possibly converge uniformly at the same time to the same point. In fact if we substitute $n = -a_\nu$ we find that for any a_ν there exists an n such that the points q_{n+a_ν} and r_{n+a_ν} are as far apart as the distinct fixed

points q and r . In other words a complementary interval e cannot exist and $C = E$. At the same time we find:

(11) *If T is transitive and for fixed p and sequence of integers a_n the sequence of points p_{a_n} converges then all sequences q_{a_n} converge uniformly in q .*

Remark. The criterion in (10) is in a certain degenerate sense still valid in case τ is rational and even when T does change the orientation of C , if we consider as transitive any transformation for which $E = C$ if τ is irrational and for which some power of T is the identical transformation in the other cases.

VIII. If T is transitive it is topologically equivalent with a rotation. So the next question that arises is to find the equations of the continuous group of transformations of which such a transformation T is one element.

We take a fixed point p on C and two variable points q and r . We will define a point s depending on q and r ($s = f(q, r)$) such that the transformation of that continuous group transforming p into q , transforms r into s .

(12) *If p_{a_n} converges to q and p_{b_m} converges to r then $p_{a_n + b_m}$ converges to $s = f(q, r)$.*

The existence and uniqueness of $s = f(q, r)$ follows from (10). That $s = f(q, r)$ is the equation of the desired continuous group is evident if we apply the transformation S (VI). We see that if Sq and Sr are the results of a rotation of Sp over angles $2\pi\lambda$ and $2\pi\mu$, then $Sf(q, r)$ is the result of a rotation of Sp over an angle $2\pi(\lambda + \mu)$. In particular it follows that $s = f(q, r)$ depends continuously on the two points q and r together. $s = f(q, r)$ satisfies the following relations:

$$\begin{aligned} s = f(q, r) &= f(r, q); f_n(q, r) = f(q_n, r); \\ f(p, r) &= r; f(q, f(r, s)) = f(f(q, r), s). \end{aligned}$$

All these properties of $f(q, r)$ could be derived directly from (10) and (11).

IX. Condition (10) can be modified if we suppose that the curve C is in the complex plane, so that any point of C is a complex number. But for a slight change, that can easily be accounted for, (10) then says:

(13) *N and S for T to be transitive is that the transforms p_n of any point p of C form an almost periodic function on the group of addition of all integers n .**

* See J. von Neumann, "Almost periodic functions in a group I," *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 445-492, and A. Walther, "Fast-periodische Folgen," *Hamburg Abhandlungen*, vol. 6 (1928), pp. 217-234.

The considerations of section VI can then be amplified in the following way: The circle D can be considered as representing the group of rotations round its center. It contains a subgroup simply isomorphic with the group of addition of integers: the group of rotations over angles $2\pi\tau n$. If we let p_n correspond to the rotation $2\pi\tau n$ then according to VI the resulting function can be extended to a continuous function on the group D , in fact to the transformation S^{-1} transforming D into C . From the theory of almost periodic functions it follows now that the two almost periodic functions p_n on the group of integers and S^{-1} on the group of rotations D have the same Fourier sequence. If $2\pi\alpha$ represents a variable angle of rotation on D then the Fourier sequence of S^{-1} must have the form $\sum_{-\infty}^{\infty} a_k \exp(2\pi i \alpha k)$. As we find the values of α corresponding to p_n by putting $\alpha = n\tau$ the Fourier sequence of p_n will have the form $\sum_{-\infty}^{\infty} a_k \exp(2\pi i n \tau k)$, in other words:

(14) *The Fourier exponents of the almost periodic function p_n of (13) will be integer multiples of τ .*

X. The equality of the Fourier coefficients of p_n and S^{-1} can be generalized, expressed in terms of the functions $y = g(x)$ and $x = h(y)$ of V and proved in a very simple way.

We need the fact that the smallest positive remainders (τn) of τn modulo one are uniformly everywhere dense between 0 and 1. So if $t(y)$ is a Riemann integrable function defined for $0 \leq y \leq 1$ then $\int_0^1 t(y) dy = \lim_{n \rightarrow \infty} (1/n) \sum_{k=1}^n t((k\tau))$. If $t(x)$ is everywhere defined, Riemann integrable and has the period one this becomes:

$$\int_0^1 t(y) dy = \lim_{n \rightarrow \infty} (1/n) \sum_{l=1}^n t(l\tau).$$

Writing $t(g(x)) = s(x)$ we find that $s(x)$ is integrable with respect to $g(x)$ in the Riemann-Stieltjes sense, $s(x+1) = s(x) + 1$ and $t(y) = s(h(y))$. Furthermore for a fixed value x' of x : $t(l\tau) = s(x'_l)$. So finally:

$$(14) \quad \lim_{n \rightarrow \infty} (1/n) \sum_{l=1}^n s(x'_l) = \int_0^1 s(x) dg(x) = \int_0^1 s(h(y)) dy.$$

Under the condition that $s(x+1) = s(x) + 1$ and that either one of the two integrals exists. In the transitive case the equality of the Fourier coefficients follows if we substitute:

$$s(x) = r(x) \exp[2\pi i n g(x)], \quad s(x'_1) = r(x'_1) \exp[2\pi i n l \tau],$$

where $r(x)$ is the (periodic) function locating C in the complex plane.

XI. The condition (10) does not give a direct connection between the transformation T or the function $f(x)$ determining T , and the transitivity or intransitivity of T . The problem of finding such a condition (at least without further restrictions) seems to be extremely difficult. In the same class are all problems of the type: What can be said about the real function character of $y = g(x)$ and $x = h(y)$ (IV) if something is known about the character of $f(x)$.

The difficulty of this problem is indicated by the following statement of which we omit the simple proof:

If two segments e and e' on C have no point in common and a topological transformation T' of e into e' is given that does not change the direction, then a transformation T of C can be constructed such that $T'p = Tp$ if p is in e , T belongs to an arbitrary τ , and (except when $\tau = 0$) T is either transitive or intransitive.

This points out that, unless strong restrictions are made on the class of functions considered, no condition, that determines the local character of $f(x)$, can have an influence on the transitivity of T or even on the value of τ .

In the next section we will give a proof of the only result available along the lines indicated in this section, the following theorem of Denjoy:

(15) *If the function $f(x)$ determining T has a continuous derivative $f'(x) \neq 0$ and of bounded variation in the segment $0 \leq x \leq 1$ then T is transitive.*

XII. We take on C a fixed point p and a segment e with p as endpoint. Then we determine a positive integer n such that either p_n or p_{-n} is the only point p_k , $|k| \leq n$ in the interior of e . We can of course find arbitrarily high such integers n by taking e sufficiently small. We only consider the case that p_{-n} is the point contained in e , then:

(16) *The two finite sequences:*

$$\begin{aligned} p_0, p_1, \dots, p_{n-1}; \\ p_{-n}, p_{1-n}, \dots, p_{-1}, \end{aligned}$$

*alternate on the curve C .**

We must show that if $0 \leq k < n$, no point of the two sequences is in the

* The simplified proof of this statement here given was also found by R. H. Fox.

interior of the segment $p_k p_{k-n}$ that has the same orientation as $p_0 p_{-n}$. We consider two cases:

a. p_l is in the segment $p_k p_{k-n}$ and $n > l \geq k - n$. Then $T^{-k} p_l$ is in $T^{-k}(p_k p_{k-n})$ or p_{l-k} in $p_0 p_{-n}$, which has been excluded for $n > l - k \geq -n$.

b. p_l is in $p_k p_{k-n}$ and $-n \leq l < k - n < 0$. Then p_k is in the segment $p_{l+n} p_l$, so p_{k-l-n} is in $p_0 p_{-n}$ and this has again been excluded for $0 < k - l - n < -l \leq n$.

From the assumptions of (15) we can conclude that $F(x) = \log f'(x)$ is continuous, has the period one and is of bounded variation in the segment $0 \leq x \leq 1$. If x is a number corresponding to p we can conclude from (16) that for some constant V :

$$|F(x) + F(x_1) + \cdots + F(x_{n-1}) - (F(x_n) + F(x_{1-n}) + \cdots + F(x_1))| < V.$$

Substituting the definition of $F(x)$ and contracting the logarithms we find:

$$\left| \log \left(\frac{dx_n}{dx} \frac{dx_{-n}}{dx} \right) \right| < V,$$

$$e^{-V} < \frac{dx_n}{dx} \frac{dx_{-n}}{dx} < e^{+V}.$$

Now we take a segment e on C . We denote its length by μ , the length of $T^n e$ by μ_n and an integral over a segment of the real axis corresponding to e

by $\int_e dx$. Then:

$$\mu_n = \int \frac{dx_n}{dx} dx, \quad \mu_{-n} = \int_e \frac{dx_{-n}}{dx} dx;$$

$$\mu_n + \mu_{-n} = \int_e \left(\frac{dx_n}{dx} + \frac{dx_{-n}}{dx} \right) dx \geq 2 \int_e \left(\frac{dx_n}{dx} \frac{dx_{-n}}{dx} \right)^{1/2} dx > 2\mu e^{-V/2};$$

$$\sum_{-n}^{+n} \mu_k > 2n\mu e^{-V/2}.$$

As the sum of the measures of the segments $T^n e$ is not bounded some of these segments must overlap on C and as this is not the case for segments of $C - E$ such segments cannot exist and T must be transitive.

One consequence is that T is transitive if $f(x)$ and its inverse function are analytic.

XIII. The difficulties found in trying to conclude anything about $g(x)$

from the real function character of $f(x)$, disappears completely if we start from the function $s = f(q, r)$.

We suppose that T is transitive, so that the functions $g(x)$ and $h(y)$ are both continuous, and monotonic increasing. Then we define a function $f(x, y)$ of two variables by the equation:

$$(17) \quad f(x, y) = h[g(x) + g(y)] \text{ or } gf(x, y) = g(x) + g(y).$$

It is quite obvious that $f(x, y)$ is continuous and has the period 1 in both variables. Its meaning on the curve C is given by the function $f(q, r)$ of (12) if we take there for p the point with coördinate zero.

- (18) a. If $f(x, y)$ has the derivative $f_x(x, y)$ then the derivatives $g'(x)$ and $h'(x)$ of $g(x)$ and $h(x)$ exist and are positive.
 b. If $f_x(x, y)$ is continuous in y then $g'(x)$ and $h'(x)$ are continuous.
 c. If $f_x(x, y)$ has an n -th derivative to y , then $g^{(n+1)}(x)$ exists, $n > 0$.
 d. If $f(x, y)$ is analytic then $g(x)$ and $h(x)$ are analytic.

a. As $g(x)$ is continuous and monotonic increasing it has certainly a derivative for some value z of x . For any given x we can find a value y such that $z = f(x, y)$. For the values of x and y so determined we take the derivative of $gf(x, y) = g(x) + g(y)$ to x and find that $g(x)$ has a derivative for all values of x . Then the formula:

$$(19) \quad g'f(x, y) \cdot f_x(x, y) = g'(x),$$

now valid for all x and y shows that if $g'(x)$ was zero for one particular value of x it would always be zero and that is obviously impossible. So $g'(x)$ is positive, the inverse function $h(x)$ has a positive derivative and $f(x, y)$ has a positive total derivative.

b. If $f_x(x, y)$ is continuous in y then (19) shows, taking x fixed and $f(x, y) = z$ arbitrary, that $1/g'(x)$ and $h'(z)$ are continuous. As $1/g'(x)$ can never be equal to zero $g'(z)$ is also continuous. Finally we see from (19) that $f_x(x, y)$ has a positive greatest lower bound.

c. If $f_x(x, y)$ has an n -th derivative to y , $n > 0$, then it is continuous in y , so $f_x(x, y) > \epsilon > 0$. Now we can successively take n times the derivative of (19) to y , each time proving that the next derivative of $g(x)$ exists. In fact each time it occurs in exactly one term and with a positive coefficient.

d. If $f(x, y)$ is analytic then $f_y(x, y)$ is continuous and positive. So for

constant $z = f(x, y)$ we can solve for y as an analytic function of x . Then (19) shows that $g'(x)$ is an analytic function of x . The same is then true for $g(x)$ and $h(x)$.

LITERATURE.

- a. H. Poincaré, *Oeuvres complètes*, t. 1, pp. 137-158.
- b. P. Bohl, *Acta Mathematica*, vol. 40 (1916), pp. 321-336.
- c. H. Kneser, "Reguläre Kurvenscharen auf den Ringflächen," *Mathematische Annalen*, vol. 91 (1924), pp. 135-154.
- d. J. Nielsen, "On topologiske Afbildninger . . .," *Matematisk Tidsskrift*, B. (1928), pp. 39-46.
- e. A. Denjoy, *Comptes Rendus*, vol. 194 (1932), pp. 830-833 and pp. 2014-2016.
- f. A. Denjoy, "Sur les courbes définies par les équations différentielles à la surface du tore," *Journal de Mathématique* (9), t. 11 (1932), pp. 333-375.
- g. H. Kneser, "Periodische Differentialgleichungen . . .," *Annali di Matematica*, vol. 11 (1933), pp. 181-185.
- h. D. C. Lewis and A. Wintner, "On the distribution of successive images in the Poincaré transformation problem of a circle into itself," *American Journal of Mathematics*, vol. 56 (1934), pp. 407-410.

THE JOHNS HOPKINS UNIVERSITY.

A THEORY OF POSITIVE INTEGERS IN FORMAL LOGIC.*

PART I.

By S. C. KLEENE.

1. **Introduction.** In this paper we shall be concerned primarily with the development of the system of logic based on a set of postulates proposed by A. Church.† Our object is to demonstrate empirically that the system is adequate for the theory of positive integers, by exhibiting a construction of a significant portion of the theory within the system. By carrying out the construction on the basis of a certain subset of Church's formal axioms, we show that this portion at least of the theory of positive integers can be deduced from logic without the use of the notions of *negation*, *class*, and *description*.

Instead of limiting our discussion to the system of Church, we shall employ his rules of procedure in a generalized form, and present our results as valid for a logic based on these rules and any set of well-formed formal axioms which includes 1, 3-11, 14-16.‡ This program is explained in the introductory section of a previous article.§

We presuppose familiarity with the contents of the first article of Church, and of §§ 1-5 of our previous article.

As has been noted, significances can be assigned to the symbols of the logic in such a way that the formulas become assertions of logical truths. The mathematical interest arises from the fact that among the logical entities of the system it is possible to select certain ones which occur in relations of the same form as the relations between certain entities in mathematical theories. Hence if the mathematical entities are identified with, or defined to be, the logical entities, the propositions in which they occur will read as theorems of mathematics. It is in this sense that we are to develop or deduce the theory of positive integers within the logic, *i. e.* we are to define the numbers and

* Received October 9, 1933. Revised manuscript received June 18, 1934.

† A. Church, "A set of postulates for the foundation of logic," *Annals of Mathematics*, vol. 33 (1932), pp. 346-366, and a second paper under the same title, vol. 34 (1933), pp. 839-864. We shall refer to these articles by their dates.

‡ Church, 1933, p. 841, or 1932, p. 356.

§ S. C. Kleene, "Proof by cases in formal logic," *Annals of Mathematics*, vol. 35 (1934), no. 3. References to theorems or sections of this paper will be made by prefixing the letter *C* to the number of the theorem or section.

other notions employed in the theory as expressions of the logic, and prove formulas which assert, in the symbolism of the logic, that these expressions stand in the relations which the mathematical theory requires.

2. Equality. The definition

$$= \rightarrow \lambda\mu\nu \cdot \phi(\mu) \supset_{\phi} \phi(\nu),$$

$\{=\}$ (\mathbf{x}, \mathbf{y}) abbreviated to $[\mathbf{x}] = [\mathbf{y}]$, and the theorems *

$$2.1: \quad \mathbf{x} \cdot \mathbf{x} = \mathbf{x},$$

$$2.2: \quad [x = y] \supset_{xy} y = x,$$

$$2.3: \quad [x = y] \supset_{xy} [y = z] \supset_z x = z,$$

show that we may carry out within the system certain familiar operations with equalities.

Specifically, it follows from the definition that we may pass from an expression \mathbf{J} to an expression \mathbf{J}' by the substitution for a free occurrence of \mathbf{N} in \mathbf{J} of an equal expression \mathbf{N}' , provided the resulting occurrence of \mathbf{N}' in \mathbf{J}' is also free, *i. e.* under these circumstances $\mathbf{J}, \mathbf{N} = \mathbf{N}' \vdash \mathbf{J}'$. For then, according to C5I, we may convert \mathbf{J} into $\{\lambda\mathbf{x} \cdot \mathbf{K}\}(\mathbf{N})$ and \mathbf{J}' into $\{\lambda\mathbf{x} \cdot \mathbf{K}\}(\mathbf{N}')$ for a suitably chosen \mathbf{x} and \mathbf{K} , and we can pass from $\{\lambda\mathbf{x} \cdot \mathbf{K}\}(\mathbf{N})$ to $\{\lambda\mathbf{x} \cdot \mathbf{K}\}(\mathbf{N}')$ by means of Rule V using $\mathbf{N} = \mathbf{N}'$ as major premise. This argument also applies directly to the substitution of occurrences of \mathbf{N}' for each of a set of occurrences of \mathbf{N} , provided the occurrences of \mathbf{N} and of \mathbf{N}' are free. (Cf. C2IX.) As a special case of this substitution rule we may pass from $\mathbf{A} = \mathbf{B}$ to $\mathbf{A} = \mathbf{C}$ where \mathbf{C} is obtained by legitimate substitutions of \mathbf{N}' for \mathbf{N} within \mathbf{B} .

From 2.1 it follows that we may equate an expression \mathbf{A} to itself, or to any expression \mathbf{B} obtained from \mathbf{A} by conversion, provided only that we can prove $E(\mathbf{A})$ or $E(\mathbf{B})$. For if we have $E(\mathbf{A})$ we infer $\mathbf{A} = \mathbf{A}$ by 2.1, and convert the \mathbf{A} on the right into \mathbf{B} . If we have $E(\mathbf{B})$ we infer $\mathbf{B} = \mathbf{B}$ by 2.1, and convert the \mathbf{B} on the left into \mathbf{A} . (Cf. § C5, paragraph 2.) C5II characterizes the situations in which we may prove $E(\mathbf{M})$.

2.2 enables us to pass from $\mathbf{A} = \mathbf{B}$ to $\mathbf{B} = \mathbf{A}$, and 2.3 from $\mathbf{A} = \mathbf{B}$ and $\mathbf{B} = \mathbf{C}$ to $\mathbf{A} = \mathbf{C}$.

The abbreviation $\mathbf{A}_1 = \mathbf{A}_2 = \mathbf{A}_3 = \dots = \mathbf{A}_n$ will be used for $[\mathbf{A}_1 = \mathbf{A}_2]$

* Church, 1933, Theorems 1, 2, 3.

$[A_2 = A_3] [A_3 = A_4] \cdots [A_{n-1} = A_n]$. As a consequence of 2.2 and 2.3, each of the n^2 equalities $A_i = A_j$ ($i, j = 1, \cdots, n$) is a consequence of the chain of equalities $A_1 = A_2 = \cdots = A_n$. Frequently we are interested only in the equality $A_1 = A_n$.

We shall adopt the practice of starting with an expression, A_1 , and writing successively $= A_2, = A_3, \cdots, = A_n$, whenever it can be shown that, for an arbitrary F , $F(A_1) \vdash F(A_2)$ and $F(A_2) \vdash F(A_1)$, $F(A_2) \vdash F(A_3)$ and $F(A_3) \vdash F(A_2)$, \cdots , $F(A_{n-1}) \vdash F(A_n)$ and $F(A_n) \vdash F(A_{n-1})$, respectively. Such a chain of contingent equalities linking A_1, A_2, \cdots, A_n represents a chain of provable equalities if $E(A_i)$ is provable for any one i ($1 \leq i \leq n$). For then each of $E(A_1), E(A_2), \cdots, E(A_{n-1})$ follows from $E(A_i)$ and the facts $F(A_1) \vdash F(A_2), \cdots, F(A_n) \vdash F(A_{n-1})$; and then each of the $n-1$ factors of $A_1 = A_2 = \cdots = A_n$ can be proved with the aid of Theorem I. A special case in which $E(A_i)$ can be proved is that $A_{i-1} = A_i$ or $A_i = A_{i+1}$ is a previously established equality.

The circumstances which we most commonly use to justify linking A_{j+1} to A_j in forming a chain of contingent equalities, *i. e.* circumstances which imply that $F(A_j) \vdash F(A_{j+1})$ and $F(A_{j+1}) \vdash F(A_j)$, are the following: (1) A_j convertible into A_{j+1} ,* (2) $A_j = A_{j+1}$ or $A_{j+1} = A_j$ previously established, (3) A_{j+1} obtainable from A_j by a substitution of N' for N of the type described above ($N = N'$ or $N' = N$ being previously established).†

Ordinarily we introduce a chain of contingent equalities only as a preliminary to inferring their provability, or their provability from formulas which we have assumed. It is intended that it should be clear from the context when this is the case. It will suffice to point out in the course of building the chain or subsequently that one of the equalities is already known, or that one of the expressions exists. Even this formality may be omitted in situations of a type which has already occurred often, when we evidently make subsequent use of the chain of equalities as proved or as proved under our assumptions.

However chains of contingent equalities are of interest in themselves since a chain of contingent equalities linking N and N' can be used in place of a proved $N = N'$ in passing from J to J' under the conditions above, *i. e.* $J \vdash J'$ can be inferred by means of the former. In particular, in the construction of one chain of contingent equalities we may use, (2'), instead of $A_i = A_{i+1}$ or $A_{i+1} = A_i$ under (2), a previously obtained chain of contingent equalities

* In this case we may write *conv*, instead of $=$, before A_{j+1} in constructing the chain.

† (3) includes (2). The transitive property, $A = B, B = C \vdash A = C$, of equality is a special case of the substitution rule, $J, N = N' \vdash J'$.

linking A_i and A_{i+1} , and, (3'), instead of $N = N'$ or $N' = N$ under (3), a previously obtained chain linking N and N' . Chains of contingent equalities are derived in § 15, under hypotheses which do not imply their provability, with a view to their subsequent use in this manner.

If one chain of contingent equalities is used in obtaining another in the fashion just described, and the latter established as true, then the provability of the former follows. For $E(A_i)$ or $E(N)$ can then be proved according to C5II. Thus starting with conversions and known equalities we can build a hierarchy of contingent equalities. If one of the expressions linked in the chain at the top occurs in a provable formula as a free expression, then all the equalities of the hierarchy are provable.

As with formulas, contingent equalities $M = N$ will often occur below when it is shown merely that they hold as a consequence of "assumptions" X, Y, Z, \dots , i. e. when $F(M), X, Y, Z, \dots \vdash F(N)$ and $F(N), X, Y, Z, \dots \vdash F(M)$.

When there is ambiguity, we may use an accent to distinguish a contingent equality $M = 'N$ from an equality $M = N$. The assertion that $M = 'N$ would hold if X, Y, Z, \dots were added to the list of axioms may be written $X, Y, Z, \dots \vdash 'M = 'N$.

3. Definition of the positive integers. Peano's axioms. A construction of the theory of positive integers in the logic has been begun by Church,* who selected formulas to represent the three undefined terms, one, successor, and the notion of being a number, of Peano,† and adopted Peano's definitions of 2, 3, \dots by means of the first two of them, thus:

$$\begin{aligned} 1 &\rightarrow \lambda f x \cdot f(x). \\ S &\rightarrow \lambda p f x \cdot f(p(f, x)). \\ N &\rightarrow \lambda \mu \cdot [\phi(1) \cdot \phi(x) \supset_x \phi(S(x))] \supset_\phi \phi(\mu). \\ 2 &\rightarrow S(1), \quad 3 \rightarrow S(2), \quad 4 \rightarrow S(3), \dots \end{aligned}$$

We follow Church in these definitions, but not in the definition of addition, because J. B. Rosser has proposed one which leads to simpler formal proofs, nor in the definitions of multiplication and subtraction, because it is our program to avoid the use of the function ι .

Our first objective will, of course, be the formulation and proof in the logic of Peano's axioms, except the fourth which involves negation. Formula-

* Church, 1933, § 9.

† G. Peano, *Rivista di Matematica*, vol. 1 (1891), pp. 87-102. Peano called the "first" number "zero."

tions have been given by Church, which we adopt except in the case of the fifth. The fifth we express by 3. 3, which may be a better rendering of Peano's axiom, and which we need in the course of our development of the theory.

The first two, 3. 1 and 3. 2, may be proved thus:

- \mathfrak{A}_1 : $\Sigma(\Sigma)$ —by two applications of IV to any axiom.
 \mathfrak{A}_2 : $\{\lambda\rho \cdot \rho(I, \Sigma(\Sigma))\}(1)$ —by conversion from \mathfrak{A}_1 . ($I \rightarrow \lambda x \cdot x$)
 \mathfrak{A}_3 : $\Sigma y \cdot \{\lambda\rho \cdot \rho(I, \Sigma(\Sigma))\}(y)$ —IV, (III, \mathfrak{A}_2).
 \mathfrak{D}_y : $\{\lambda\rho \cdot \rho(I, \Sigma(\Sigma))\}(y)$ —assumed, in preparation for an application of Theorem I.
 \mathfrak{C}_{y1} : $\{\lambda\rho \cdot \rho(I, \Sigma(\Sigma))\}(S(y))$ —by conversion from \mathfrak{D}_y .
 \mathfrak{A}_4 : $\{\lambda\rho \cdot \rho(I, \Sigma(\Sigma))\}(y) \supset_y \{\lambda\rho \cdot \rho(I, \Sigma(\Sigma))\}(S(y))$ —provable, according to Theorem I, using \mathfrak{A}_3 and the proof of \mathfrak{C}_{y1} from \mathfrak{D}_y .
 \mathfrak{A}_5 : $\Sigma\phi \cdot \phi(1) \cdot \phi(y) \supset_y \phi(S(y))$ —IV, III, (14, \mathfrak{A}_2 , \mathfrak{A}_4).
 \mathfrak{D}_ϕ : $\phi(1) \cdot \phi(y) \supset_y \phi(S(y))$ —assumed.
 $\mathfrak{C}_{\phi 2}$: $\phi(1)$ —15, \mathfrak{D}_ϕ .
 \mathfrak{A}_6 : $[\phi(1) \cdot \phi(y) \supset_y \phi(S(y))] \supset_\phi \phi(1)$ —Theorem I, \mathfrak{A}_5 , $\mathfrak{C}_{\phi 2}$.
3. 1: $N(1)$ —III, \mathfrak{A}_6 .
 \mathfrak{A}_7 : $\Sigma x \cdot N(x)$ —IV, (III, 3. 1).
 \mathfrak{D}_x : $N(x)$ —assumed.
 $\mathfrak{C}_{\phi x 3}$: $\phi(x)$ —V, \mathfrak{D}_x , \mathfrak{D}_ϕ , conversion.
 $\mathfrak{C}_{\phi 4}$: $\phi(y) \supset_y \phi(S(y))$ —16, \mathfrak{D}_ϕ .
 $\mathfrak{C}_{\phi x 5}$: $\phi(S(x))$ —V, $\mathfrak{C}_{\phi 4}$, $\mathfrak{C}_{\phi x 3}$, conversion.
 $\mathfrak{C}_{x 6}$: $[\phi(1) \cdot \phi(y) \supset_y \phi(S(y))] \supset_\phi \phi(S(x))$ —Theorem I, \mathfrak{A}_5 , $\mathfrak{C}_{\phi x 5}$.
 $\mathfrak{C}_{x 7}$: $N(S(x))$ —III, $\mathfrak{C}_{x 6}$.
3. 2: $N(x) \supset_x N(S(x))$ —Theorem I, \mathfrak{A}_7 , $\mathfrak{C}_{x 7}$.

The fifth Peano axiom we establish as follows:

- \mathfrak{A}_1 : $N(y)N(y) \supset_y N(S(y))$ —provable by Theorem I, using 3. 1, 3. 2, 14, 15.
 \mathfrak{D}_ϕ : $\phi(1) \cdot N(y)\phi(y) \supset_y \phi(S(y))$ —assumed; $\Sigma\phi \cdot \mathfrak{D}_\phi$ is provable from 14, 3. 1, \mathfrak{A}_1 .
 $\mathfrak{C}_{\phi 1}$: $N(1)\phi(1)$ —14, 3. 1, 15(\mathfrak{D}_ϕ).
 $\mathfrak{D}_{\phi, y}$: $N(y)\phi(y)$ —assumed; $\Sigma y \cdot \mathfrak{D}_{\phi, y}$ is provable by means of III and IV from $\mathfrak{C}_{\phi 1}$.
 $\mathfrak{C}_{\phi y 2}$: $\phi(S(y))$ —provable by means of 16(\mathfrak{D}_ϕ) and $\mathfrak{D}_{\phi, y}$.
 $\mathfrak{C}_{y 3}$: $N(S(y))$ —provable by means of 3. 2, 15($\mathfrak{D}_{\phi, y}$).
 $\mathfrak{C}_{\phi 4}$: $N(y)\phi(y) \supset_y N(S(y))\phi(S(y))$ —Theorem I, 14($\mathfrak{C}_{\phi y 2}$, $\mathfrak{C}_{y 3}$).
 $\mathfrak{C}_{\phi 5}$: $N(1)\phi(1) \cdot N(y)\phi(y) \supset_y N(S(y))\phi(S(y))$ —14, $\mathfrak{C}_{\phi 1}$, $\mathfrak{C}_{\phi 4}$.
 \mathfrak{D}_x : $N(x)$ —assumed; $\Sigma x \cdot N(x)$ is provable by means of 3. 1.

$\mathfrak{G}_{\phi x_6}$: $N(x)\phi(x) \text{---} V, \mathfrak{D}_x, \mathfrak{G}_{\phi_5}, \text{conversion.}$

$\mathfrak{G}_{\phi x_7}$: $\phi(x) \text{---} 16, \mathfrak{G}_{\phi x_6}.$

3.3: $[\phi(1) \cdot N(y)\phi(y) \supset_y \phi(S(y))] \supset_\phi \cdot N(x) \supset_x \phi(x) \text{--- Theorem I, (Theorem I, } \mathfrak{G}_{\phi x_7}).$

The third Peano axiom will be proved as Theorem 10.1 below.

4. Proof by induction. In the development of the theory of positive integers by means of intuitive logic, the object of Peano's fifth axiom is to justify proofs by induction. By following Frege in making it the definition of x being a positive integer that propositions involving p for which a mathematical induction with respect to p can be carried out should hold when p is taken to be x , we are enabled to express the fifth axiom as a provable formula of the logic. If the logic is adequate, it should then be possible to carry out proofs by induction within the logic. This we show to be the case.

4I. *If x is a variable which does not occur in F as a free variable, if $\vdash F(1)$, and if $N(x), F(x) \vdash F(S(x))$, then $\vdash N(x) \supset_x F(x)$.*

For by means of 14, 3.1, and $F(1)$, we can prove $N(1)F(1)$, and thence, since x does not occur in F as a free symbol and, being a variable, is distinct from the free symbols, Π , $\&$, of N , we can obtain $\{\lambda x \cdot N(x)F(x)\}(1)$ by conversion. An application of IV gives $\Sigma x \cdot N(x)F(x)$, and, assuming $N(x)F(x)$, we can prove $N(x)$ and $F(x)$ by 15 and 16, and thence, according to the third hypothesis of the Theorem, $F(S(x))$. Then by Theorem I: $N(x)F(x) \supset_x F(S(x))$. Combining this with $F(1)$ by 14, we obtain a formula convertible into $\{\lambda \phi \cdot \phi(1) \cdot N(y)\phi(y) \supset_y \phi(S(y))\}(F)$ since $F, N, \&, S$ do not contain x as a free symbol. Using this formula as minor premise with 3.3 as major premise to an application of Rule V, we obtain $\{\lambda \phi \cdot N(x) \supset_x \phi(x)\}(F)$, and by conversion, since F does not contain x as a free variable, $N(x) \supset_x F(x)$. Under these circumstances we may say that $N(x) \supset_x F(x)$ is proved by *induction with respect to x* from the *basis $F(1)$* ; and we may call $F(x)$ the *hypothesis of the induction*. This terminology, with appropriate modifications, will also be used in connection with the generalizations of this theorem.

Certain generalizations of the simple inductive procedure described in 4I are permissible, because they can be reduced to one or more simple inductions. Although the nature of these reductions is quite evident intuitively, we shall give them explicitly to ensure that they can be carried out wholly within the confines of the system.

4II. If x is a variable which does not occur in F_1, F_2, \dots, F_n as a free variable, if $\vdash F_1(1), \vdash F_2(1), \dots, \vdash F_n(1)$, and if $N(x), F_1(x), F_2(x), \dots, F_n(x) \vdash F_i(S(x))$ ($i=1, \dots, n$), then $\vdash N(x) \supset_x F_1(x), \vdash N(x) \supset_x F_2(x), \dots, \vdash N(x) \supset_x F_n(x)$.

For then we can take $\lambda x \cdot F_1(x) F_2(x) \dots F_n(x)$ as the F of the simple case, and prove $N(x) \supset_x F_1(x) F_2(x) \dots F_n(x)$.^{*} With the aid of this result we can prove each of the theorems $N(x) \supset_x F_1(x), \dots, N(x) \supset_x F_n(x)$ by means of Theorem I. The restrictions on x have been used tacitly in this argument. A set of theorems which are inferred to be provable by an application of this theorem may be said to be proved by a *simultaneous induction*.

4III. If x is a variable which does not occur in F as a free variable, if $\vdash F(1), \vdash F(2), \dots, \vdash F(n)$, and if $N(x), F(x), F(S(x)), \dots, F(S(\dots n-1 \text{ times} \dots (S(x)) \dots)) \vdash F(S(\dots n \text{ times} \dots (S(x)) \dots))$, then $\vdash N(x) \supset_x F(x)$.[†]

This theorem we establish by an intuitive induction with respect to n . It has been proved as 4I for the case that n is 1. We assume it for a value p of n , and apply it under the hypotheses with n taken as $p+1$ to the function $\lambda x \cdot F(x) F(S(x))$, obtaining $N(x) \supset_x F(x) F(S(x))$. Thence by Theorem I and Axiom 15 we can prove $N(x) \supset_x F(x)$. Under the circumstances of this theorem we may say that $N(x) \supset_x F(x)$ is proved by induction with respect to x from the n -tuple basis $F(1), F(2), \dots, F(n)$.

4IV. If x_1, x_2, \dots, x_n are distinct variables which do not occur as free variables in F , and if $\vdash F(1, 1, \dots, 1); N(x_1), F(x_1, 1, \dots, 1) \vdash F(S(x_1), 1, \dots, 1); N(x_1), N(x_2), F(x_1, x_2, \dots, 1) \vdash F(x_1, S(x_2), \dots, 1); \dots; N(x_1), N(x_2), \dots, N(x_n), F(x_1, x_2, \dots, x_n) \vdash F(x_1, x_2, \dots, S(x_n))$, then $\vdash N(x_1) N(x_2) \dots N(x_n) \supset_{x_1 x_2 \dots x_n} F(x_1, x_2, \dots, x_n)$.

This theorem concerning *induction with respect to n variables* reduces to 4I in case n is 1. If the theorem be assumed for the case that n is p , and applied under the hypotheses with n taken as $p+1$ to the function $\lambda x_1 x_2 \dots x_{n-1} \cdot F(x_1, \dots, x_{n-1}, 1)$, we obtain $N(x_1) \dots N(x_{n-1}) \supset_{x_1 \dots x_{n-1}}$

^{*} Henceforth we often omit mention of applications of Axioms 14-16.

[†] When n is being used to represent a given positive integer of intuitive logic, we may use n to represent the corresponding positive integer $S(\dots n-1 \text{ times} \dots (S(1)) \dots)$ of the formal logic, and *vice versa*. We sometimes employ symbols of the formal theory in a familiar intuitive sense, the context being supposed to indicate when this is being done. For example, 1, n , $+$, $-$ in $x_1, x_n, x_{n+1}, n-1 \text{ times}$, and n -th.

$\cdot F(x_1, \dots, x_{n-1}, 1)$. Also by means of the last hypothesis and a corollary of Theorem I we obtain that $N(x_n), N(x_1) \cdot \dots \cdot N(x_{n-1}) \supset_{x_1 \dots x_{n-1}} \cdot F(x_1, \dots, x_{n-1}, x_n) \vdash N(x_1) \cdot \dots \cdot N(x_{n-1}) \supset_{x_1 \dots x_{n-1}} F(x_1, \dots, x_{n-1}, S(x_n))$. Hence, applying 4I to the function $\lambda x_n \cdot N(x_1) \cdot \dots \cdot N(x_{n-1}) \supset_{x_1 \dots x_{n-1}} \cdot F(x_1, \dots, x_n)$, we can prove $N(x_n) \supset_{x_n} N(x_1) \cdot \dots \cdot N(x_{n-1}) \supset_{x_1 \dots x_{n-1}} \cdot F(x_1, \dots, x_n)$, and thence, by a corollary of Theorem I, $N(x_1) \cdot \dots \cdot N(x_n) \supset_{x_1 \dots x_n} F(x_1, \dots, x_n)$.

Finally we may vary the simple inductive process of 4I in a combination of the three directions represented by 4II, 4III, and 4IV, and justify the procedure by a corresponding combination of the devices used in establishing 4II, 4III, and 4IV.

We shall apply the theorems of this section, and similar theorems below, when the situation exhibited is not exactly that described by the theorem, but one that could be made so by evident conversions. (Cf. § C5.)

5. Addition. We adopt the definition

$$+ \rightarrow \lambda \rho \sigma f x \cdot \rho(f, \sigma(f, x)),$$

due to J. B. Rosser, and abbreviate $\{+\}(x, y)$ to $[x] + [y]$.

5I. $\{+\}(1) \text{ conv } S. [x + y] + z \text{ conv } x + [y + z]. S(x + y) \text{ conv } S(x) + y.$

The last conversion follows from the first two thus: $S(x + y) \text{ conv } 1 + x + y \text{ conv } [1 + x] + y \text{ conv } S(x) + y$. It is often more convenient to employ this theorem and like theorems of §§ 6, 7 (often tacitly) than to refer to the formal theorems proved by means of them.

Since $E(S)$ can be proved (*e. g.* from 3.2), the first conversion of 5I leads, by means of 2.1,* to

$$5.1: \quad \{+\}(1) = S.$$

Assume $N(y)$. Then, (1) we can prove $N(S(y))$ by means of 3.2, and thence, by conversion, $N(1 + y)$, and (2) if we assume $N(x + y)$, we can prove $N(S(x + y))$ by 3.2, and thence, by conversion, $N(S(x) + y)$. Having (1) and (2), we can prove $N(x) \supset_x N(x + y)$ by induction.† This was done on the assumption $N(y)$, and $\Sigma y \cdot N(y)$ is provable from 3.1.

* Cf. the remarks on 2.1 in § 2.

† Cf. 4I. In this case only the second of the assumptions $N(x), F(x)$ is used.

By Theorem I we obtain $N(y) \supset_y \cdot N(x) \supset_x N(x+y)$, and thence, by a corollary of Theorem I,*

$$5.2: \quad N(x)N(y) \supset_{xy} N(x+y).$$

5.2 with 15 and 16 enable us to prove $E([x+y] + z)$ as a consequence of $N(x)N(y)N(z)$, and hence the second conversion of 5I leads to

$$5.3: \quad N(x)N(y)N(z) \supset_{xyz} \cdot [x+y] + z = x + [y+z].$$

We shall use $x + y + z$ as an abbreviation for $[x + y] + z$.

(1) Since $E(2)$ is provable (*e.g.* from 3.1, 3.2) and 2 is convertible into $1 + 1$, we can prove $1 + 1 = 1 + 1$. (2) Assume $N(x)$ and $x + 1 = 1 + x$. Now $S(x) + 1 \text{ conv } S(x+1), = S(1+x)$ (by means of the assumption $x + 1 = 1 + x$), $\text{conv } S(S(x))$ (and $E(S(S(x)))$) is provable from $N(x)$ with the aid of 3.2), $\text{conv } 1 + S(x)$. Hence, by § 2, $S(x) + 1 = 1 + S(x)$ is provable from our assumptions. (3) Using (1) and (2) we can prove, by induction

$$\mathfrak{A}_1: \quad N(x) \supset_x \cdot x + 1 = 1 + x.$$

(4) Assume $N(x)$, $N(y)$, and $x + y = y + x$. Then $x + S(y) \text{ conv } x + \cdot 1 + y, = x + \cdot y + 1$ (since $y + 1 = 1 + y$ is provable from $N(y)$ and \mathfrak{A}_1), $\text{conv } [x+y] + 1, = 1 + [x+y]$ (by means of \mathfrak{A}_1 , $N(x)$, $N(y)$, 5.2—this equality is proved from our assumptions), $= 1 + [y+x]$ (by means of the hypothesis $x + y = y + x$), $\text{conv } S(y) + x$. This establishes $x + S(y) = S(y) + x$ from our assumptions, according to § 2. Having (1), (2), and (4), we infer by induction (*cf.* 4IV)

$$5.4: \quad N(x)N(y) \supset_{xy} \cdot x + y = y + x.$$

Another fundamental theorem on addition will be obtained as 11.4 below.

6. Multiplication. We adopt the definition,

$$\times \rightarrow \lambda \rho \sigma x \cdot \rho(\sigma(x)),$$

of J. B. Rosser, and abbreviate $\{\times\}(x, y)$ to $[x] \times [y]$ or $[x][y]$.†

* In making applications of Theorem I and its corollaries we may omit mention of the formulas $\Sigma x \cdot M$, $\Sigma xy \cdot M$, . . . , when their proof is obvious, as here from 3.1 and 14. (Cf. the statements of Theorem I and its corollaries, Church, 1932, pp. 358, 366.)

† The abbreviation $[x][y]$ for the product of two positive integers, x and y , employed in the presence of $N(x)$ and $N(y)$, should not lead to confusion with the

6I. $[xy]z \text{ conv } x[yz]. [x + y]z \text{ conv } xz + yz.$

(1) $3.1 \vdash E(1).$ $E(1), 2.1 \vdash 1 = 1$. By conversion from $1 = 1$, $[1 \cdot 1] = 1$. (2) Assume $N(x)$. Then $E(S(x))$ is provable by means of 3.2. $1S(x) \text{ conv } S(x)$. Hence $[1S(x)] = S(x)$. From (1) and (2) by induction: *

$$6.1: \quad N(x) \supset_x \cdot 1x = x.$$

Assume $N(y)$. Then (1) by means of 6.1 we have $N(1y)$, and (2) assuming $N(x)$ and $N(xy)$, we can prove $N(1y + xy)$ by means of 5.2, $N(xy)$, and (1), and pass by conversion to $N(S(x)y)$. Having (1) and (2) we infer by induction that $N(y) \vdash N(x) \supset_x N(xy)$. With this result we can prove, by Theorem I Corollary,

$$6.2: \quad N(x)N(y) \supset_{xy} N(xy).$$

Assuming $N(x)N(y)N(z)$, we can prove $E([xy]z)$ by means of 6.2, and $E([x + y]z)$ by means of 5.2 and 6.2. Hence 6I leads to the pair of formal theorems

$$6.3: \quad N(x)N(y)N(z) \supset_{xyz} \cdot [xy]z = x[yz],$$

$$6.4: \quad N(x)N(y)N(z) \supset_{xyz} \cdot [x + y]z = xz + yz.$$

Assume $N(k)$. Then: (1) $1[1 + k] = [1 \cdot 1] + 1k$, by 3.2, 2.1, conversion and 6.1. (2) Assume also $N(l)$ and $l[1 + k] = l1 + lk$. Then $S(l)[1 + k] \text{ conv } 1[1 + k] + l[1 + k], = 1 + k + l[1 + k]$ (by 6.1, 3.1, 5.2), $= 1 + k + l1 + lk$ (by our assumption $l[1 + k] = l1 + lk$), $\text{conv } 1 + [k + l1] + lk, = 1 + [l1 + k] + lk$ (by 3.1, 6.2, 5.4), $\text{conv } [1 \cdot 1] + l1 + k + lk, = [1 \cdot 1] + l1 + 1k + lk$ (by 6.1), $\text{conv } S(l)1 + S(l)k$. We can prove the existence of any one of these expressions from our assumptions by means of 3.1, 5.2, 6.2. Hence, by § 2, $S(l)[1 + k] = S(l)1 + S(l)k$. (3) By induction, using (1) and (2), $N(l) \supset_l \cdot l[1 + k] = l1 + lk$. According to Theorem I, this argument enables us to infer the provability of

$$\mathfrak{N}_1: \quad N(k) \supset_k \cdot N(l) \supset_l \cdot l[1 + k] = l1 + lk.$$

abbreviation for the logical product of two propositions. The abbreviation $[x][y][z]$ for $[[x][y]][z]$ will be employed with the arithmetical product as with the logical product.

* Cf. 4I. In this case only the first of the assumptions $N(x)$ and $F(x)$ is used.

(1) By conversion from $1 = 1 : [1 \cdot 1] = 1 \cdot 1$. (2) Assume $N(x)$ and $x1 = 1x$. Then $S(x)1 \text{ conv } 1 + x1, = 1 + 1x$ (by the second assumption), $= 1 + x$ (by 6.1 and $N(x)$), $\text{conv } S(x)(E(S(x)))$ is provable from $N(x)$, using 3.2), $\text{conv } 1S(x)$. Hence by § 2: $S(x)1 = 1S(x)$. (3) By induction, using (1) and (2),

$$\mathfrak{A}_2: \quad N(x) \supset_x \cdot x1 = 1x.$$

(4) Assume $N(x), N(y), xy = yx$. Then $xS(y) \text{ conv } x[1 + y], = x1 + xy$ (by $\mathfrak{A}_1, N(x), N(y)$), $= x1 + yx$ (by the hypothesis $xy = yx$), $= 1x + yx$ (by $\mathfrak{A}_2, N(x)$), $\text{conv } S(y)x$. Hence $xS(y) = S(y)x$. Having (1), (2), and (4), we can prove, by induction (cf. 4IV),

$$6.5: \quad N(x)N(y) \supset_{xy} \cdot xy = yx.$$

7. Exponentiation. $1(F, A) \text{ conv } F(A)$, and $S(x, F, A) \text{ conv } F(x(F, A))$. By an intuitive induction, utilizing the definitions of 2, 3, 4, \dots from 1, we infer that for any given positive integer z , $\mathfrak{z}(F, A)$ is convertible into $F(\dots z \text{ times } \dots (F(A)) \dots)$. Church's definitions of the positive integers were framed with a view to providing this formal means of representing the z -th power of a function F of an argument A . We recognize it by introducing the abbreviation $[x]^{[p]}$ for $\{p\}(x)$, so that $\mathfrak{z}(F, A)$ may be written $F^{\mathfrak{z}}(A)$.

3.1, 3.2, 14 $\vdash \Sigma \phi a f \cdot \phi(a) \cdot \phi(\rho) \supset_{\rho} \phi(f(\rho))$. Assuming $\phi(a) \cdot \phi(\rho) \supset_{\rho} \phi(f(\rho))$, we have, by the use of the first factor with the second, $\phi(f(a))$, which is convertible into $\phi(f^2(a))$. Assuming also $\phi(f^x(a))$, the second factor gives $\phi(f(f^x(a)))$, which is convertible into $\phi(f^{S(x)}(a))$. Hence, by induction and the corollary of Theorem I,

$$7.1: \quad [\phi(a) \cdot \phi(\rho) \supset_{\rho} \phi(f(\rho))] \supset_{\phi a f} \cdot N(x) \supset_x \phi(f^x(a)).$$

(1) $I^1 = I$ is provable from $I = I$ by conversion.* (2) Assuming $I^x = I$, we have $I^{S(x)} \text{ conv } \lambda y \cdot I^x(y), = \lambda y \cdot I(y)$ (by means of our assumption), $\text{conv } I$. $\vdash E(I)$. Hence, by § 2, $I^{S(x)} = I$. From (1) and (2) by induction,

$$7.2: \quad N(x) \supset_x \cdot I^x = I.$$

It follows, since $I(A) \text{ conv } A$, that we can express formally a function of positive integers having any given constant value A . Indeed, if

* $I = I$ is provable by 2.1 since $\vdash E(I)$ (cf. §§ C6, C8).

$$\mathfrak{G} \rightarrow \lambda\pi \cdot [\lambda\rho\sigma \cdot I^\sigma(\rho)]^\pi,$$

then $\mathfrak{G}(1, A), \mathfrak{G}(2, A), \dots$ are functions whose values for any one, two, \dots members of the sequence $1, 2, \dots$ (i. e. "known" positive integers), respectively, are convertible into A .

The formula 6.1 is convertible into

$$7.3: \quad N(x) \supset_x \cdot [\lambda fa \cdot f^x(a)] = x.$$

$$7I. \quad F^x(F^y(A)) \text{ conv } F^{x+y}(A).$$

Let f and a be proper symbols not occurring in x and y as free symbols. Then the formula $x + y$ is convertible into $\lambda fa \cdot f^x(f^y(a))$ and $\lambda fa \cdot f^{x+y}(a)$. This leads evidently to the conversion 7I, and also shows from a new point of view why the definition given in § 5 for the addition of positive integers is suitable.

The observation is due to J. B. Rosser that if p and q are positive integers the formula x^p is actually the p -th power of x in the arithmetic sense.

$$7II. \quad x^p x^q \text{ conv } x^{p+q}. \quad [x^p]^q \text{ conv } x^{qp}.$$

8. Number dyads and triads. A finite ordered set of expressions can be defined intuitively by enumerating its members, A_1, A_2, \dots, A_k . In order to carry out an argument in the logic about such a set whose members vary according to given laws, we may require a formula A such that (1) A is a function of the k members A_1, \dots, A_k , and (2) A_1, \dots, A_k are k functions of A . If there exist expressions H_1, \dots, H_k such that $H_1(A_1) = I, \dots, H_k(A_k) = I$, we can take as A the expression $\lambda x \cdot x(A_1, \dots, A_k)$, where x is any proper symbol not occurring in A_1, \dots, A_k as a free symbol. In the important case that A_1, \dots, A_k are positive integers there are the simpler constructions which follow.

To represent the ordered pair of numbers x and y , we employ the formula $D(x, y)$, abbreviated $[x, y]$, where

$$D \rightarrow \lambda\rho\sigma fga \cdot f^\rho(g^\sigma(a)).$$

Then x and y are the functions D_1 and D_2 ,

$$D_1 \rightarrow \lambda\rho f \cdot \rho(f, I) \quad \text{and} \quad D_2 \rightarrow \lambda\rho \cdot \rho(I),$$

respectively, of $[x, y]$, as is established in the theorems

$$8.1: \quad N(x)N(y) \supset_{xy} \cdot \mathbf{D}_1([x, y]) = x,$$

$$8.2: \quad N(x)N(y) \supset_{xy} \cdot \mathbf{D}_2([x, y]) = y.$$

The first is provable, according to § 2 and the corollary of Theorem I, since assuming $N(x)N(y)$ we have $\mathbf{D}_1([x, y]) \text{ conv } \lambda fa \cdot f^x(I^y(a)), = \lambda fa \cdot f^x(I(a))$ (by 7.2, $N(y)$), $\text{conv } \lambda fa \cdot f^x(a), = x$ (by 7.3, $N(x)$); the second, since $\mathbf{D}_2([x, y]) \text{ conv } \lambda fa \cdot I^x(f^y(a))$, from which we can pass similarly to y .

Let

$$\mathbf{T} \rightarrow \lambda \rho \sigma \tau f g h a \cdot f^\rho(g^\sigma(h^\tau(a))),$$

and abbreviate $\mathbf{T}(x, y, z)$ to $[x, y, z]$. If

$$\mathbf{T}_1 \rightarrow \lambda \rho f \cdot \rho(f, I, I), \quad \mathbf{T}_2 \rightarrow \lambda \rho f \cdot \rho(I, f, I), \quad \mathbf{T}_3 \rightarrow \lambda \rho \cdot \rho(I, I),$$

then

$$8.3: \quad N(x)N(y)N(z) \supset_{xyz} \cdot \mathbf{T}_1([x, y, z]) = x,$$

$$8.4: \quad N(x)N(y)N(z) \supset_{xyz} \cdot \mathbf{T}_2([x, y, z]) = y,$$

$$8.5: \quad N(x)N(y)N(z) \supset_{xyz} \cdot \mathbf{T}_3([x, y, z]) = z.$$

Similarly for tetrads, etc.

9. Predecessor. Given the third Peano axiom, as formulated in 10.1, a function for the predecessor of a positive integer can be obtained by means of the description, ι . Conversely, given a predecessor function, 10.1 can be proved. By obtaining a predecessor function first, and making it the basis of the proof of 10.1, and of further definitions, we are able to limit our formal assumptions, avoiding, particularly, the use of the description.

The predecessor function, P , which we shall define has, besides the property that if x is one of the expressions $1, 2, \dots$, then $P(S(x)) \text{ conv } x$, also the property $P(1) \text{ conv } 1$. The arithmetical functions, such as subtraction and division, defined by means of it, are likewise given a value within the class of positive integers for every set of positive integral values of the arguments, even in cases when this is not ordinarily done. The resulting special properties of the functions are useful in the development of the theory.

Let

$$\mathfrak{F} \rightarrow \lambda \rho \cdot [\mathbf{T}_2(\rho), \mathbf{T}_3(\rho), S(\mathbf{T}_3(\rho))] \quad \text{and} \quad \mathfrak{S} \rightarrow [1, 1, 1].$$

Then (1) $\mathfrak{F}^{S(1)}(\mathfrak{S}) \text{ conv } [1, 2, 3]$, and $E([1, 2, 3])$ is provable since

$[1, 2, 3]$ occurs in the proposition $T_1([1, 2, 3]) = 1$ which is provable from 8.3, 3.1, and 3.2. Hence, by 2.1, $\mathfrak{F}^{S(1)}(\mathfrak{F}) = [1, S(1), S(S(1))]$. (2) Assume $N(x)$ and $\mathfrak{F}^{S(x)}(\mathfrak{F}) = [x, S(x), S(S(x))]$. From $N(x)$ we infer $N(S(x))$, $N(S(S(x)))$, and $N(S(S(S(x))))$ by 3.2. Now $\mathfrak{F}^{S(S(x))}(\mathfrak{F}) \text{ conv } [T_2(\mathfrak{F}^{S(x)}(\mathfrak{F})), T_3(\mathfrak{F}^{S(x)}(\mathfrak{F})), S(T_3(\mathfrak{F}^{S(x)}(\mathfrak{F})))], = [T_2([x, S(x), S(S(x))]), T_3([x, S(x), S(S(x))]), S(T_3([x, S(x), S(S(x))]))]$ (by the hypothesis of the induction), $= [S(x), S(S(x)), S(S(S(x)))]$ (by means of the equalities obtained by using $N(x)$, $N(S(x))$, and $N(S(S(x)))$ in 8.4 and 8.5). $E([S(x), S(S(x)), S(S(S(x)))])$ is given by 8.3, $N(S(x))$, $N(S(S(x)))$, and $N(S(S(S(x))))$. Hence, by § 2, $\mathfrak{F}^{S(S(x))}(\mathfrak{F}) = [S(x), S(S(x)), S(S(S(x)))]$. Having (1) and (2), we can prove by induction

$$\mathfrak{A}_1: \quad N(x) \supset_x \cdot \mathfrak{F}^{S(x)}(\mathfrak{F}) = [x, S(x), S(S(x))].$$

Let

$$P \rightarrow \lambda\rho \cdot T_1(\mathfrak{F}^\rho(\mathfrak{F})).$$

Assume $N(x)$. By 3.2, $N(S(x))$ and $N(S(S(x)))$. Then $P(S(x)) \text{ conv } T_1(\mathfrak{F}^{S(x)}(\mathfrak{F})), = T_1([x, S(x), S(S(x))])$ (by $N(x)$ and \mathfrak{A}_1), $= x$ (by means of 8.3, $N(x)$, $N(S(x))$, and $N(S(S(x)))$). Hence, by § 2 and Theorem I,

$$9.1: \quad N(x) \supset_x \cdot P(S(x)) = x.$$

The theorem

$$9.2: \quad P(1) = 1$$

is provable by conversion from $1 = 1$; and the theorem

$$9.3: \quad N(x) \supset_x N(P(x))$$

is provable by induction, since (1) $N(P(1))$ follows by conversion from 3.1, and (2) assuming $N(x)$, we can obtain $P(S(x)) = x$ from 9.1, and then $N(P(S(x)))$ from $N(x)$.

10. Peano's third axiom. $N(1)N(1) \cdot S(1) = S(1)$ is provable by means of 3.1, 3.2, 2.1, and 14; and thence $\exists xy \cdot N(x)N(y) \cdot S(x) = S(y)$ can be proved. Assume $N(x)N(y) \cdot S(x) = S(y)$. Then $x = P(S(x))$ (by 9.1 and $N(x)$), $= P(S(y))$ (by $S(x) = S(y)$), $= y$ (by 9.1 and $N(y)$). Hence, by § 2 and Theorem I Cor.,

$$10.1: \quad [N(x)N(y) \cdot S(x) = S(y)] \supset_{xy} \cdot x = y.$$

11. Subtraction. Let

$$\longrightarrow \lambda\mu\nu \cdot P^\nu(\mu),$$

and abbreviate $\{\longrightarrow\}(x, y)$ to $[x] - [y]$. If x and y are positive integers, then $x - y$ has the usual significance if x is greater than y , and $x - y$ is 1 if x is less than or equal to y .

$N(x)$, 9.3, 7.1 $\vdash N(y) \supset_y N(x - y)$. Hence, by Theorem I Cor.,

$$11.1: \quad N(x)N(y) \supset_{xy} N(x - y).$$

(1) Assuming $N(x)$, we have $[x + 1] - 1 = [1 + x] - 1$ (by 5.4, $N(x)$, and 3.1), $\text{conv } P(S(x))$, $= x$ (by 9.1 and $N(x)$). Hence, by § 2 and Theorem I, $N(x) \supset_x \cdot [x + 1] - 1 = x$. (2) Assume $N(y)$ and $N(x) \supset_x \cdot [x + y] - y = x$. Assume also $N(x)$. Then $[x + S(y)] - S(y) = [S(y) + x] - S(y)$ (by 5.4, $N(y)$, 3.2, $N(x)$), $\text{conv } P([1 + y + x] - y)$, $= P([y + 1 + x] - y)$ (by 5.4, $N(y)$, 3.1), $\text{conv } P([y + S(x)] - y)$, $= P([S(x) + y] - y)$ (by 5.4, $N(y)$, $N(x)$, 3.2), $= P(S(x))$ (by the second assumption, $N(x)$, 3.2), $= x$ (by 9.1, $N(x)$). Hence, by § 2 and Theorem I, $N(x) \supset_x \cdot [x + S(y)] - S(y) = x$. From (1) and (2) by induction, $N(y) \supset_y \cdot N(x) \supset_x \cdot [x + y] - y = x$; whence, by Theorem I Cor.,

$$11.2: \quad N(x)N(y) \supset_{xy} \cdot [x + y] - y = x.$$

(1) Assume $N(x)N(y)$. Then $[x - y] - 1 \text{ conv } P(y(P, x))$, $\text{conv } S(y, P, x)$, $\text{conv } x - S(y)$, $\text{conv } x - \cdot 1 + y$, $= x - \cdot y + 1$ (by 5.4, 3.1, $N(y)$). $N(x)$, $N(y)$, 3.1, 5.2, 11.1 $\vdash N(x - \cdot y + 1)$, from which $E(x - \cdot y + 1)$ is provable. Hence, by § 2 and Theorem I Cor., $N(x)N(y) \supset_{xy} \cdot [x - y] - 1 = x - \cdot y + 1$. (2) Assume $N(z)$ and $N(x)N(y) \supset_{xy} \cdot [x - y] - z = x - \cdot y + z$. Assume also $N(x)N(y)$. Then $[x - y] - S(z) \text{ conv } [x - y] - \cdot 1 + z$, $= [[x - y] - 1] - z$ (by the hypothesis of the induction, $N(x)$, $N(y)$, 11.1, 3.1), $= [x - \cdot y + 1] - z$ (by (1), $N(x)$, $N(y)$), $= x - \cdot y + 1 + z$ (by the hypothesis of the induction, $N(x)$, $N(y)$, 3.1, 5.2), $\text{conv } x - \cdot y + S(z)$. We can prove $E(x - \cdot y + S(z))$ by means of $N(x)$, $N(y)$, $N(z)$, 3.2, 5.2, 11.1. Hence, by § 2 and Theorem I Cor., $N(x)N(y) \supset_{xy} \cdot [x - y] - S(z) = x - \cdot y + S(z)$. Having (1) and (2), we infer by induction $N(z) \supset_z \cdot N(x)N(y) \supset_{xy} \cdot [x - y] - z = x - \cdot y + z$, and thence

$$11.3: \quad N(x)N(y)N(z) \supset_{xyz} \cdot [x - y] - z = x - [y + z].$$

Assume $N(x)N(y)N(z) \cdot x + y = x + z$. Then $y = [x + y] - x$ (by 11.2, 5.4, $N(y), N(x)$), $= [x + z] - x$ (by $x + y = x + z$), $= z$ (by 11.2, 5.4, $N(x), N(z)$). Hence

$$11.4: \quad [N(x)N(y)N(z) \cdot x + y = x + z] \supset_{xyz} \cdot y = z.$$

12. Order. Let $< \rightarrow \lambda\mu\nu \cdot N(\mu) \cdot [\phi(S(\mu)) \cdot \phi(\xi) \supset_{\xi} \phi(S(\xi))] \supset_{\phi} \phi(\nu)$, and abbreviate $\{<\}(x, y)$ to $x < y$ or $y > x$.

$$12.1: \quad N(x) \supset_x \cdot x < S(x).$$

Proof. Assume $N(x)$. Then, using 3.2, $\Sigma\phi \cdot \phi(S(x)) \cdot \phi(\xi) \supset_{\xi} \phi(S(\xi)) \cdot \phi(S(x)) \cdot \phi(\xi) \supset_{\xi} \phi(S(\xi)) \vdash \phi(S(x))$. By Theorem I, $[\phi(S(x)) \cdot \phi(\xi) \supset_{\xi} \phi(S(\xi))] \supset_{\phi} \phi(S(x))$, which with $N(x)$ yields $x < S(x)$.*

$$12.2: \quad x < y \supset_{xy} \cdot N(x)N(y).$$

Proof. 12.1 $\vdash \Sigma xy \cdot x < y$. $x < y \vdash N(x) \cdot [\phi(S(x)) \cdot \phi(\xi) \supset_{\xi} \phi(S(\xi))] \supset_{\phi} \phi(y)$. $N(x), 3.2 \vdash N(S(x))$. $N(S(x)), 3.2, [\phi(S(x)) \cdot \phi(\xi) \supset_{\xi} \phi(S(\xi))] \supset_{\phi} \phi(y) \vdash N(y)$.

$$12.3: \quad [x < y \cdot y < z] \supset_{xyz} \cdot x < z.$$

Proof. Assume $x < y \cdot y < z$ and $\phi(S(x)) \cdot \phi(\xi) \supset_{\xi} \phi(S(\xi))$. The latter with $x < y$ yields $\phi(y)$. $\phi(y), \phi(\xi) \supset_{\xi} \phi(S(\xi)) \vdash \phi(S(y)) \cdot \phi(\xi) \supset_{\xi} \phi(S(\xi))$, which with $y < z$ yields $\phi(z)$. Hence $[\phi(S(x)) \cdot \phi(\xi) \supset_{\xi} \phi(S(\xi))] \supset_{\phi} \phi(z)$, and $x < z$.

When each a_i ($i = 1, \dots, n$) is either $=$ or $<$ ($=$ or $>$), we abbreviate $[x_1 a_1 x_2][x_2 a_2 x_3] \dots [x_n a_n x_{n+1}]$ to $x_1 a_1 x_2 a_2 x_3 \dots x_n a_n x_{n+1}$. By § 2 and 12.3, we can make formally the usual inferences concerning the order relation of two expressions linked in such a chain of inequalities.†

* Henceforth our proofs will be abbreviated by omissions of such details as applications of Theorem I and Corollaries, applications of § 2 and like principles for inequalities, and references to formal theorems under circumstances in which it is clear what theorems are being used. In particular, required formulas of the form $N(A)$ will not be mentioned, when they are obtainable from the hypotheses and such theorems as 3.1, 3.2, 5.2, 12.2. In case several theorems are used at a given step in the argument, those playing a subordinate rôle may not be cited.

† We arrange the introduction and proof of chains of inequalities in the same manner as that of chains of equalities. Any link may be a conversion or contingent equality.

$$12.4: \quad N(x) \supset_x \cdot N(y) \supset_y \cdot x + y > y.$$

Proof. Assume $N(x)$, $N(y) \supset_y \cdot x + y > y$, and $N(y)$. Then $S(x) + y$ conv $S(x + y)$, $> x + y$ (12.1), $> y$ (by means of the hypotheses). Hence, by Theorem I, $N(x)$, $N(y) \supset_y \cdot x + y > y \vdash N(y) \supset_y \cdot S(x) + y > y$. Thus 12.4 is provable by induction from 12.1 as basis.

$$12.5: \quad x < y \supset_{xy} \cdot y = [y - x] + x.$$

Proof. Assume $x < y$. (1) $S(x)$ conv $1 + x$, $= [[1 + x] - x] + x$ (11.2), conv $[S(x) - x] + x$. Hence $N(S(x)) \cdot S(x) = [S(x) - x] + x$. (2) Assume $N(\xi) \cdot \xi = [\xi - x] + x$. Then $S(\xi) = S([\xi - x] + x)$ (by the hyp.), conv $S(\xi - x) + x$, $= [[S(\xi - x) + x] - x] + x$ (11.2), conv $[S([\xi - x] + x) - x] + x$, $= [S(\xi) - x] + x$ (by the hyp.). Hence $N(S(\xi)) \cdot S(\xi) = [S(\xi) - x] + x$. By Theorem I, $[N(\xi) \cdot \xi = [\xi - x] + x] \supset_{\xi} \cdot N(S(\xi)) \cdot S(\xi) = [S(\xi) - x] + x$. (3) $x < y \vdash [\phi(S(x)) \cdot \phi(\xi) \supset_{\xi} \phi(S(\xi))] \supset_{\phi} \phi(y)$, which with (1) and (2) yields $N(y) \cdot y = [y - x] + x$.

$$12.6: \quad x < y \supset_{xy} \cdot x = y - [y - x].$$

$$12.7: \quad [N(x) \cdot y > z] \supset_{xyz} \cdot [x + y] - z = x + [y - z].$$

Proofs. Assuming $x < y$ and letting $\mathbf{p} \rightarrow y - x$, $x = [\mathbf{p} + x] - \mathbf{p}$ (11.2), $= [\mathbf{p} + x] - [\mathbf{p} + x] - x$ (11.2), $= y - y - x$ (12.5). Assuming $N(x) \cdot y > z$ and letting $\mathbf{p} \rightarrow y - z$, $[x + y] - z = [x + \mathbf{p} + z] - z$ (12.5), $= x + \mathbf{p}$ (11.2), $= x + [\mathbf{p} + z] - z$ (11.2), $= x + y - z$ (12.5).

$$12.8: \quad N(x) \supset_x \cdot S(x) > 1.$$

$$12.9: \quad N(y) \supset_y \cdot [x > y] \supset_x \cdot x > 1.$$

$$12.10: \quad N(x)N(y) \supset_{xy} \cdot x + y > 1.$$

Proofs. 12.4 \vdash 12.8, and 12.9 is provable by induction, since $[x > 1] \supset_x \cdot x > 1$ and $N(y) \vdash [x > S(y)] \supset_x \cdot x > 1$ (12.4). 12.10 follows from 12.9 by 12.4.

$$12.11: \quad [N(x) \cdot y < z] \supset_{xyz} \cdot x + y < x + z.$$

$$12.12: \quad [N(x)N(y)N(z) \cdot x + y < x + z] \supset_{xyz} \cdot y < z.$$

Proofs. Assuming $N(x) \cdot y < z$, then $x + z = x + y + z - y$ (12.5),

$> x + y$ (12.4). Assuming $N(x)N(y)N(z) \cdot x + y < x + z$, then $z = [x + z] - x$ (11.2), $= [p + x + y] - x$ (by 12.5, if $p \rightarrow [x + z]$ $- x + y$), $= p + [x + y] - x$ (12.7, 12.4), $= p + y$ (11.2), $> y$ (12.4).

$$12.13: \quad [N(z) \cdot x < S(y) \cdot y < S(z)] \supset_{xyz} \cdot x < S(z).$$

$$12.14: \quad [N(z) \cdot x < y \cdot y < S(z)] \supset_{xyz} \cdot x < z.$$

$$12.15: \quad [x < S(y) \cdot y < z] \supset_{xyz} \cdot x < z.$$

Proofs. Assuming $N(z) \cdot x < S(y) \cdot y < S(z)$, we have $1 + S(z) = 1 + [S(z) - y] + y$ (12.5), $= [S(z) - y] + S(y)$, $= [S(z) - y] + [S(y) - x] + x$ (12.5), $> 1 + x$ (12.11, 12.10), and hence, by 12.12, $S(z) > x$. Similarly for 12.14 and 12.15.

We let $x \leq y \rightarrow x < S(y)$,* and employ the relation \leq as well as $<$ and $=$ in our chains of inequalities. 12.13-12.15 together with previously noted facts show that we can make formally the usual inferences concerning the order relation of two expressions linked in such a chain.

$$12.16: \quad [N(y) \cdot x < S^2(y)] \supset_{xy} \cdot x - y = 1.$$

Proof. Assume $N(y) \cdot x < S^2(y)$. By 11.2, $[S^2(y) - 1] - y = 1$. Assuming $N(p)$ and $[S^2(y) - p] - y = 1$, $[S^2(y) - S(p)] - y = [[S^2(y) - p] - y] - 1$ (11.3), $= 1 - 1$ (by the hyp.), conv 1. Hence, by induction, $N(p) \supset_p \cdot [S^2(y) - p] - y = 1$. This with $N(S^2(y) - x)$ yields $[S^2(y) - S^2(y) - x] - y = 1$, and, by 12.6, $x - y = 1$.

$$12.17: \quad [x < S(y) \cdot y < S(x)] \supset_{xy} \cdot x = y.$$

Proof. Assume $x < S(y) \cdot y < S(x)$. $x < S(y)$, 12.11 $\vdash S(x) < S^2(y)$. (1) $2 + y$ conv $S^2(y)$, $= [S^2(y) - S(x)] + S(x)$ (12.5, $S(x) < S^2(y)$), $= [S^2(y) - S(x)] + [S(x) - y] + y$ (12.5, $y < S(x)$). Hence, by 11.4, (2) $2 = [S^2(y) - S(x)] + S(x) - y$. (3) $S^2(y) - S(x) = [[S^2(y) - S(x)] + S(x) - y] - S(x) - y$ (11.2), $= 2 - S(x) - y$ (by (2)), $= 2 - 1$ (12.16, $S(x) < S^2(y)$), conv 1. (4) $2 + x$ conv $1 + S(x)$, $= [S^2(y) - S(x)] + S(x)$ (by (3)), $= 2 + y$ (as in (1)). Hence, by 11.4, $x = y$.

$$12.18: \quad [x > 1] \supset_x \cdot N(y) \supset_y \cdot x - y < x.$$

* Or let $\leq \rightarrow \lambda xy \cdot x < S(y)$ and abbreviate $\{\leq\}(x, y)$ to $x \leq y$. Similarly below..

Proof. Assume $x > 1$. (1) $1 - 1 \text{ conv } 1, < S(1)$. Assuming $N(p)$, $S(p) - 1 = p, < S^2(p)$. Hence, by induction, $N(p) \supset_p \cdot p - 1 < S(p)$. (2) $x - 1 < S(x - 1), = x$ (12.5). (3) Assuming $N(y)$ and $x - y < x$, $x - S(y) = [x - y] - 1$ (11.3), $\leq x - y$ (by means of (1)), $< x$ (hyp. induction). From (2) and (3), by induction, $N(y) \supset_y \cdot x - y < x$.

12.19: $N(k) \supset_k \cdot [\rho < S(k) \supset_\rho \phi(\rho)] [\rho > k \supset_\rho \phi(\rho)] \supset_\phi \cdot N(x) \supset_x \phi(x)$.

Proof. Let $\mathfrak{D}_\phi(\mathbf{a}) \rightarrow [\rho < S(\mathbf{a}) \supset_\rho \phi(\rho)] [\rho > \mathbf{a} \supset_\rho \phi(\rho)]$. (1) $\vdash \Sigma \phi \cdot \mathfrak{D}_\phi(1)$. By induction and Theorem I, $\mathfrak{D}_\phi(1) \supset_\phi \cdot N(x) \supset_x \phi(x)$. (2) Assume $N(k)$ and $\mathfrak{D}_\phi(k) \supset_\phi \cdot N(x) \supset_x \phi(x)$. Then $\Sigma \phi \cdot \mathfrak{D}_\phi(S(k))$. Assume $\mathfrak{D}_\phi(S(k))$. (a) $\rho < S(k) \vdash \rho < S^2(k)$; and hence, by the first factor of $\mathfrak{D}_\phi(S(k))$, and Theorem I, $\rho < S(k) \supset_\rho \phi(\rho)$. (b) $N(k) \vdash 1 + k < S^2(k)$, whence, by the first factor of $\mathfrak{D}_\phi(S(k))$, $\phi(1 + k)$. Assuming $N(p)$, we have $S(p) + k > S(k)$, whence, by the second factor of $\mathfrak{D}_\phi(S(k))$, $\phi(S(p) + k)$. By induction, $N(p) \supset_p \phi(p + k)$. Thence, assuming $\rho > k$, we obtain $\phi([\rho - k] + k)$, and, by 12.5, $\phi(\rho)$. By Theorem I, $\rho > k \supset_\rho \phi(\rho)$. (c) $\mathfrak{D}_\phi(k) \supset_\phi \cdot N(x) \supset_x \phi(x)$ with (a) and (b) yields $N(x) \supset_x \phi(x)$. By Theorem I, $\mathfrak{D}_\phi(S(k)) \supset_\phi \cdot N(x) \supset_x \phi(x)$. (3) 12.19 follows from (1) and (2) by induction.

13. The lesser and greater of two positive integers. Let $\min \rightarrow \lambda xy \cdot S(y) - \cdot S(y) - x$ and $\max \rightarrow \lambda xy \cdot [x + y] - \min(x, y)$.

- 13.1: $N(x)N(y) \supset_{xy} N(\min(x, y))$.
 13.2: $[N(y) \cdot x < S(y)] \supset_{xy} \cdot \min(x, y) = x$.
 13.3: $N(x) \supset_x \cdot N(y) \supset_y \cdot \min(x, y) = \min(y, x)$.
 13.4: $N(x)N(y) \supset_{xy} \cdot \min(x, y) < S(y)$.
 13.5: $N(x)N(y)N(z) \supset_{xyz} \cdot \min(z + x, z + y) = z + \min(x, y)$.

Proofs. 3.2, 11.1 \vdash 13.1. 12.6 \vdash 13.2. 13.3 may be established by an application of 12.19, since, assuming $N(x)$, (1) assuming $y < S(x)$, $\min(x, y) \text{ conv } S(y) - \cdot S(y) - x, = S(y) - 1$ (by 12.16, since $y < S(x)$, 12.11 $\vdash S(y) < S^2(x)$), $= y, = \min(y, x)$ (13.2), and (2) assuming $y > x$, a like series of steps takes us from $\min(y, x)$ to $\min(x, y)$. 12.18, 12.8 \vdash 13.4. \vdash 13.5, since, assuming $N(x)N(y)N(z)$, $\min(z + x, z + y) \text{ conv } S(z + y) - \cdot S(z + y) - z + x, = [z + S(y)] - \cdot S(y) - x$ (11.2, 11.3), $= z + \cdot S(y) - \cdot S(y) - x$ (12.7, 12.18, 12.8), $\text{conv } z + \min(x, y)$.

14. Proof by cases. We now establish theorems which connect the present theory with that of §§ C7, C9, C10. $M \rightarrow \lambda \mu \cdot \phi(1)\phi(2) \supset_\phi \phi(\mu)$,

and $[P] \equiv_{x_1 \dots x_n} Q$ will be used as an abbreviation for $[[P] \supset_{x_1 \dots x_n} Q]$
 $[[Q] \supset_{x_1 \dots x_n} P]^*$

14.1: $M(x) \equiv_x x < 3$.

Proof. C7.1 $\vdash \Sigma x \cdot M(x)$. $1 < 3$, $2 < 3$, $M(x) \vdash x < 3$, by Rule V. Hence $M(x) \supset_x x < 3$. Conversely, assume $x < 3$ and $\phi(1)\phi(2)$. Then $\phi(\min(1, 2))$ ($\phi(1)$, 13.2); and assuming $N(p)$, $\phi(\min(S(p), 2))$ ($\phi(2)$, 13.2, 13.3, 12.4). Hence, by induction, $N(p) \supset_p \phi(\min(p, 2))$. Thence, since $x < 3 \vdash N(x)$, we obtain $\phi(\min(x, 2))$, and, by 13.2, $\phi(x)$.

Note that $M(x) \vdash N(x)$ (by 14.1, 15).

Let $[x] \circ [y] \rightarrow \min(x, y)$. \circ multiplies 1's and 2's as 0's and 1's resp.

- 14.2: $M(x)M(y) \supset_{xy} M(x \circ y)$.
 14.3: $M(x)M(y) \supset_{xy} x \circ y = y \circ x$.
 14.4: $M(y) \supset_y 1 \circ y = 1$.
 14.5: $M(y) \supset_y 2 \circ y = y$.
 14.6: $[M(x)M(y) \cdot x \circ y = 2] \equiv_{xy} [x = 2] [y = 2]$.

Proofs. 14.5 follows from $2 \circ 1 = 1$ and $2 \circ 2 = 2$ by the definition of M . For 14.6a (i. e. the first factor of 14.6), assume $M(x)M(y) \cdot x \circ y = 2$. Then, by 13.4, $x \circ y < S(y)$. Hence $2 < S(y)$. Also, by 14.1, $y < S(2)$. Hence, by 12.17, $y = 2$. Likewise $x = 2$ (cf. 14.3).

Let $\epsilon \rightarrow \lambda xy \cdot \min(2, S(x) - y)$, and abbreviate $\epsilon(x, y)$ to ϵ_y^x .

- 14.7: $N(x)N(y) \supset_{xy} M(\epsilon_y^x)$.
 14.8: $[N(y) \cdot x < S(y)] \equiv_{xy} N(x)N(y) \cdot \epsilon_y^x = 1$.
 14.9: $x > y \equiv_{xy} N(x)N(y) \cdot \epsilon_y^x = 2$.

Proofs. 13.3, 13.4, 14.1 \vdash 14.7. 12.16, 12.11 \vdash 14.8a; and for 14.9a we have, assuming $x > y$, $\epsilon_y^x = \min(2, S(x) - y)$, $= \min(2, S(x - y))$ (12.7), $= 2$ (by 13.2, since $12.4 \vdash S^2(x - y) > 2$). 14.8b and 14.9b we prove as follows: By C7I, there exists a formula \mathfrak{B} such that $\mathfrak{B}(i)$ conv $\lambda ab \cdot a < S(b)$ and $\mathfrak{B}(2)$ conv $\lambda ab \cdot a > b$. Assuming $N(y)$, then (1) assuming $x < S(y)$ we infer $\epsilon_y^x = 1$ by 14.8a, and hence $\mathfrak{B}(\epsilon_y^x, x, y)$, and (2) assuming $x > y$, we infer $\epsilon_y^x = 2$ by 14.9a, and hence $\mathfrak{B}(\epsilon_y^x, x, y)$. By 12.19 and Theorem I, $N(y) \supset_y N(x) \supset_x \mathfrak{B}(\epsilon_y^x, x, y)$. This lemma enables us to infer $x < S(y)$ from $N(x)N(y) \cdot \epsilon_y^x = 1$ and $x > y$ from $N(x)N(y) \cdot \epsilon_y^x = 2$.

* Cf. Church, 1932, p. 355.

Let $\delta \rightarrow \lambda xy \cdot 4 \rightarrow \epsilon_y^x + \epsilon_x^y$, and abbreviate $\delta(x, y)$ to δ_y^x .

- 14.10: $N(x)N(y) \supset_{xy} M(\delta_y^x)$.
 14.11: $N(x)N(y) \supset_{xy} \delta_y^x = \delta_x^y$.
 14.12: $x < y \supset_{xy} \delta_y^x = 1$.
 14.13: $[N(y) \cdot x < S(y) \cdot \delta_y^x = 1] \supset_{xy} x < y$.
 14.14: $[N(x)N(y) \cdot x = y] \equiv_{xy} N(x)N(y) \cdot \delta_y^x = 2$.

Proofs. Assuming $N(x)N(y)$, $\delta_y^x = 4 \rightarrow \epsilon_y^x + \epsilon_x^y$, $= [4 \rightarrow \epsilon_y^x] \rightarrow \epsilon_x^y$ (11.3), $< 4 \rightarrow \epsilon_y^x$ (by 12.18, since $4 \rightarrow \epsilon_y^x = 1 + \cdot 3 \rightarrow \epsilon_y^x$ (12.7, 14.1, 14.7), > 1 (12.10)), ≤ 3 (12.18), and hence, by 14.1, $M(\delta_y^x)$. 5.4 \vdash 14.11. Assuming $x < y$, $\delta_y^x = 4 \rightarrow \epsilon_y^x + 2$ (14.9), $= [4 \rightarrow 2] \rightarrow \epsilon_y^x$ (11.3), conv $2 \rightarrow \epsilon_y^x$, $= 1$ (12.16). Assuming $N(y) \cdot x < S(y) \cdot \delta_y^x = 1$, $\epsilon_x^y = 3 \rightarrow \cdot 3 \rightarrow \epsilon_x^y$ (12.6, 14.1, 14.7), $= 3 \rightarrow \cdot 4 \rightarrow \cdot 1 + \epsilon_x^y$ (11.2, 11.3), $= 3 \rightarrow \cdot 4 \rightarrow \cdot \epsilon_y^x + \epsilon_x^y$ (14.8), $= 3 \rightarrow \delta_y^x$ (def.), $= 3 \rightarrow 1$ (hyp.), conv 2; and hence, by 14.9, $x < y$. Assuming $N(x)N(y) \cdot x = y$, then $x < S(y)$ and, by 14.8, $\epsilon_y^x = 1$; also $\epsilon_x^y = 1$; hence $\delta_y^x = 4 \rightarrow \cdot 1 + 1$, conv 2. It remains to establish 14.14b, which we do as follows. Assume $N(x)N(y) \cdot \delta_y^x = 2$. By 14.7, $M(\epsilon_y^x)$. (1) Assuming $\epsilon_y^x = 1$, then $x < S(y)$ (14.8). (2) Assuming $\epsilon_y^x = 2$, then $x > y$ (14.9); hence $\delta_y^x = 1$ (14.11, 14.12). Hence, by C10II, $x < S(y)$. Similarly, $y < S(x)$. Hence, by 12.17, $x = y$.

Henceforth we use tacitly results of this section in conjunction with C9I and C10II. The headings "Case 1," "Subcase a," etc., will indicate applications of C9I or C10II.

PRINCETON UNIVERSITY.

ON THE COOLING OF THE EARTH.

By ARNOLD N. LOWAN.

ABSTRACT.

In an earlier paper * the thermal history of the earth was investigated under the assumption that the earth is a radioactive sphere, the surface of which is permanently at 0° . The present paper introduces the further refinement of taking into account the heat evolved as a result of the contraction accompanying the cooling of the earth. The method is essentially that based on the use of the Laplace transformation as presented in the earlier paper above mentioned in conjunction with the theory of biorthogonal sets of functions and the corresponding integral equations with asymmetric kernels.

The theory underlying the subsequent derivations is due to the work of J. M. C. Duhamel † who nearly a hundred years ago has set up the differential equations governing the cooling of a solid when account is taken of the heat set free as a result of its accompanying contraction. As an illustration of his theory, Duhamel has obtained the complete solution for the problem of the cooling of a sphere (with due account of contraction) the surface of which is permanently at 0° . The present paper contemplates the generalization of Duhamel's problem to the case where the sphere is radioactive.

As in A. N. L. it will be assumed that the sphere (earth) is centrally symmetrical with regard to the initial temperature distribution, the distribution of radioactive matter and that the physical constants do not vary with depth.

In view of the results in J. D. and A. N. L., it is clear that the mathematical formulation of our problem is as follows:

$$(1) \quad \frac{\partial}{\partial t} T(r, t) = k \left(\frac{\partial^2}{\partial r^2} T(r, t) + \frac{2}{r} \frac{\partial}{\partial r} T(r, t) \right) + \frac{C_p - C_v}{3C_v \delta} \frac{\partial}{\partial t} \left(3\psi + r \frac{\partial \psi}{\partial r} \right) + \phi(r, t)$$

$$(2) \quad r \frac{\partial^2}{\partial r^2} \psi(r, t) + 4 \frac{\partial}{\partial r} \psi(r, t) - \frac{5\delta}{3} \frac{\partial}{\partial r} T(r, t) = 0$$

* Arnold N. Lowan, "On the cooling of a radioactive sphere," *Physical Review*, November 1, 1933. This paper will be referred to as A. N. L.

† *Journal de l'école Polytechnique* 1837. This paper will be referred to as J. D.

$$(3) \quad \lim_{t \rightarrow 0} T(r, t) = f(r)$$

$$(4) \quad \lim_{t \rightarrow 0} \psi(r, t) = 0$$

$$(5) \quad T(R, t) = 0$$

$$(6) \quad 5\psi(r, t) + 3r(\partial/\partial r)\psi(r, t) = 0 \quad \text{for } r = R$$

where the significance of the symbols employed is as follows:

$T(r, t)$ = temperature at time t , and distance r from center;

R = radius of the earth;

k = thermal diffusivity = ratio between the thermal conductivity K and the product of the density ρ and the specific heat C_v (under constant volume);

C_p = specific heat under constant pressure;

δ = coefficient of linear thermal expansion;

$\phi(r, t) = 1/\rho C_v \times$ heat generated per unit time per unit volume by the radioactive matter;

$r\psi$ = increment of the radius vector which initially had the value r .

It may be remarked that if $C_p = C_v$, the problem becomes identical with that treated in A. N. L.; further the system (1) to (6) is identical with that treated by Duhamel, except for the term $\phi(r, t)$ in the differential equation (1). Thus the solution to be derived must reduce to that given in A. N. L. if $C_p = C_v$ and to that given by Duhamel if $\phi(r, t) = 0$.

Consider the differential equation

$$(7) \quad r \partial^2 \psi / \partial^2 + 4 \partial \psi / \partial r = 0.$$

If we assume a solution of (7) in the form $\psi = Ar^m$ it is readily found that $m(m+3) = 0$. Thus the general solution of (7) is in the form $\psi = A + Br^{-3}$. Starting with this solution of (7) the solution of (2) satisfying the boundary condition (6) and such that $r\psi$ is finite for $r=0$ is readily found by the method of variation of parameters in the form:

$$(8) \quad \psi(r) = \frac{5\delta}{r^3} \int_0^r \rho^2 T(\rho) d\rho + \frac{4\delta}{3R^3} \int_0^R \rho^2 T(\rho) d\rho$$

whence ultimately

$$(9) \quad \frac{\partial}{\partial t} \left(3\psi + r \frac{\partial \psi}{\partial r} \right) = \frac{5\delta}{3} \frac{\partial T}{\partial t} + \frac{4\delta}{3} \int_0^R \rho^2 \frac{\partial T}{\partial t} d\rho.$$

In view of (8), equation (1) becomes

$$10) \quad \frac{\partial T}{\partial t} = k \left(\frac{\partial^2 T}{\partial r^2} + \frac{2}{r} \frac{\partial T}{\partial r} \right) + \frac{5}{9} \frac{C_v - C_p}{C_v} \frac{\partial T}{\partial t} + \frac{4}{3} \frac{C_v - C_p}{R^3 C_v} \int_0^R \rho^2 T(\rho) d\rho + \phi(r, t).$$

If we make the substitution:

$$(11) \quad T(r, t) = (1/r)u(r, t) + rf(r).$$

It is ultimately found that the function $u(r, t)$ must satisfy the system of equations.

$$(12) \quad A^2 \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial r^2} + F''(r) - B^2 r \int_0^R \rho \frac{\partial u}{\partial t} d\rho + \frac{r}{k} \phi(r, t)$$

$$(13) \quad \lim_{t \rightarrow 0} u(r, t) = 0$$

$$(14-15) \quad u(0, t) = u(R, t) = 0$$

where we have put:

$$(16) \quad A^2 = \frac{1}{k} \left(1 + \frac{5}{9} \frac{C_p - C_v}{C_v} \right); \quad B^2 = \frac{4}{3} \frac{C_p - C_v}{R^3 k C_v}; \quad F(r) = rf(r).$$

In accordance with the method employed in A. N. L. we now operate on (12), (14) and (15) by the "Laplace operator" defined as follows:

$$(17) \quad L\{u(r, t)\} = \int_0^\infty e^{-\lambda t} u(r, t) dt = y(r, \lambda)$$

$$(18) \quad L\{\phi(r, t)\} = \psi(r, \lambda).$$

If we make use of the identity *

$$(19) \quad L\{(\partial/\partial t)u(r, t)\} = \lambda L\{u(r, t)\} - u(r, 0) = \lambda y - F(r)$$

it is ultimately found that the function $y(r, \lambda)$ must satisfy the equations:

$$(20) \quad y'' + \alpha^2 y = -\alpha^2 \frac{B^2}{A^2} r \int_0^R \rho y(\rho) d\rho - \sigma(r, \lambda)$$

$$(21-22) \quad y(0) = y(R) = 0$$

where we have put:

$$(23) \quad -\lambda A^2 = \alpha^2 \quad \sigma(r, \lambda) = (1/\lambda)F''(r) + (r/k)\psi(r, \lambda).$$

To solve the system (20) to (22) we start with the simpler system

$$(24) \quad y'' + \alpha^2 y = -\alpha^2 \frac{B^2}{A^2} r \int_0^R \rho y(\rho) d\rho = -\alpha^2 C r \quad (\text{say})$$

$$(20-21) \quad y(0) = y(R) = 0.$$

If we consider provisionally C as a given constant, the solution of (24), (20) and (21) is readily found in the form:

* See A. N. L.

$$(25) \quad y(r, \alpha) = C \left(-r + \frac{R}{\sin R\alpha} \sin r\alpha \right)$$

whence

$$(26) \quad \int_0^R \rho y(\rho) d\rho = -\frac{CR^3}{3} + \frac{CR}{\sin R\alpha} \cdot \frac{\sin R\alpha - R\alpha \cos R\alpha}{\alpha^2}.$$

By the definition of the constant C from (24) we have:

$$C = \frac{B^2}{A^2} \int_0^R \rho y(\rho) d\rho.$$

Therefore in view of (26) we ultimately get:

$$(27) \quad 1 - R\alpha \cot R\alpha = R^2\alpha^2 \left(\frac{1}{3} + \frac{A^2}{B^2 R^3} \right).$$

Thus the solution of the system (24), (20) and (21) is given by (25) where the α 's are the roots of the transcendental equation (27) and where now C is an arbitrary constant.

Consider now the system

$$(28) \quad z'' + \beta^2 z = 0$$

$$(29) \quad z(0) = 0$$

where the solution of (28) will be subjected to the condition of satisfying an additional boundary condition to be subsequently determined.

Let α_m be a characteristic value of the system S consisting of (24), (20) and (21) and $y_m(r)$ the corresponding characteristic function. Let similarly β_n designate a characteristic value of the system T consisting of (28), (29) and the as yet undetermined additional boundary condition and $z_n(r)$ the corresponding characteristic function. Then by the familiar "Green process" and in view of the boundary conditions (20), (21) and (29) we get

$$(30) \quad y'_m(R)z_n(R) + (\alpha_m^2 - \beta_n^2) \int_0^R y_m(\rho)z_n(\rho) d\rho \\ + \alpha_m^2 C_m \int_0^R \rho z_n(\rho) d\rho = 0$$

where C_m is the value of the constant C as defined in (24) for $y = y_m$.

It is therefore clear that if we put the condition:

$$(31) \quad y'_m(R)z_n(R) + \alpha_m^2 C_m \int_0^R \rho z_n(\rho) d\rho = 0$$

the functions $y_m(r)$ and $z_n(r)$ will satisfy the "biorthogonal" condition

$$(32) \quad \int_0^R z_n(\rho)y_m(\rho) d\rho = 0 \quad \text{if } m \neq n.$$

It will be shown that (31) is identically satisfied provided that the α_n 's and β_n 's satisfy the transcendental equation (27).

We evidently have

$$\begin{aligned} (33) \quad \int_0^R \rho z_n(\rho) d\rho &= -\frac{1}{\beta_n^2} \int_0^R \rho z_n''(\rho) d\rho \\ &= \frac{1}{\beta_n^2} \{z_n(R) - z_n(0) - Rz'_n(R)\}. \end{aligned}$$

The solution of (28) satisfying (29) is

$$(34) \quad z_n(r) = M_n \sin \beta_n r$$

where M_n is an arbitrary constant. Further from (25) we get:

$$(35) \quad y'_m(R) = C_m(-1 + R\alpha_m \cot R\alpha_m).$$

With the aid of (33), (34) and (35) the boundary condition (31) becomes

$$(36) \quad \left\{ \frac{1}{\alpha_m^2} (-1 + R\alpha_m \cot R\alpha_m) - \frac{1}{\beta_n^2} (-1 + R\beta_n \cot R\beta_n) \right\} \sin R\beta_n = 0.$$

In view of (27) it is clear that (36) and therefore (31) are identically satisfied provided the α_n 's and β_n 's are the roots of the transcendental equation (27).

From (35) and (27) we get:

$$(37) \quad \frac{y'_m(R)}{\alpha_m^2 C_m} = -R^2 \left(\frac{1}{3} + \frac{A^2}{B^2 R^3} \right) = -h \quad (\text{say}).$$

Thus (31) may be written in the more convenient form:

$$(31') \quad h z_n(R) = \int_0^R \rho z_n(\rho) d\rho.$$

This is the additional boundary condition for the function $z_n(r)$, previously referred to.

Summarizing the results found thus far the solution of the systems S and T are:

$$(25') \quad y_n(r, \alpha) = M_n \left(-r + \frac{R}{\sin R\alpha_n} \sin r\alpha_n \right)$$

$$(34') \quad z_n(r, \alpha) = N_n \sin r\alpha_n$$

where M_n and N_n are arbitrary constants and α_n is a root of the transcendental equation (27).

In the subsequent derivations it will be assumed that the solutions (25') and (34') have been normalized in accordance with:

$$(38) \quad \int_0^R y_n(r) z_n(r) dr = 1$$

whence

$$(39) \quad M_n N_n = \frac{1}{\int_0^R \sin \alpha_n \rho \left(-\rho + \frac{R}{\sin R \alpha_n} \sin \rho \alpha_n \right) d\rho}.$$

We now return to the system S and define the "Green function" $K(r, \xi)$ as the solution of the system satisfying the discontinuity condition:

$$(40) \quad \lim_{\epsilon \rightarrow 0} \frac{\partial}{\partial r} K(r, \xi) \bigg|_{\xi+\epsilon}^{\xi-\epsilon} = -1.$$

With the aid of (40) it can be readily verified that the expression:

$$(41) \quad y(r, \alpha) = \int_0^R K(r, \xi) \sigma(\xi, \lambda) d\xi$$

is the complete solution of the differential equation (20) satisfying the boundary conditions (21) and (22).

Consider now the differential equation satisfied by $K(r, \xi)$

$$(42) \quad K''(r, \xi) + \alpha^2 K(r, \xi) = -\alpha^2 \frac{B^2}{A^2} r \int_0^R \rho K(\rho, \xi) d\rho$$

in conjunction with the differential equation:

$$(28') \quad z_n'' + \beta_n^2 z_n = 0$$

satisfied by $z_n(r)$. If we perform the "Green process" on (42) and (28') and make use of the discontinuity condition (40), we ultimately get:

$$(43) \quad (\alpha^2 - \alpha_n^2) \int_0^R K(\rho, \xi) z(\rho) d\rho + z_n(\xi) = 0$$

which is an integral equation satisfied by the characteristic functions z_n .

In entirely similar fashion we define the Green function $G(r, \xi)$ as the solution of the system T satisfying the discontinuity condition

$$(44) \quad \lim_{\epsilon \rightarrow 0} \frac{\partial}{\partial r} G(r, \xi) \bigg|_{\xi+\epsilon}^{\xi-\epsilon} = -1$$

if then we perform the Green process on (42) and:

$$(45) \quad G'' + \alpha^2 G = 0$$

and take into account the boundary and discontinuity conditions satisfied by the Green functions $K(r, \xi)$ and $G(r, \xi)$ we readily find the important relation:

$$(46) \quad K(r, \xi) = G(\xi, r)$$

Furthermore by a method entirely similar to that which has led to the integral equation (43) satisfied by the characteristic functions z_n , we obtain an integral equation satisfied by the characteristic functions y in the form:

$$(47) \quad (\alpha^2 - \alpha_n^2) \int_0^R G(\rho, \xi) y(\rho) d\rho + y_n(\xi) = 0.$$

Thus the functions y_n and z_n are solutions of the integral equations (43) and (47). Since they satisfy the "biorthogonal" condition (32) they are said to form a biorthogonal set. As previously stated these functions are assumed to be normalized in accordance with (38).

The fact that the functions y_n and z_n satisfy the integral equations (43) and (47) with asymmetric kernels (in view of (46)) is of paramount importance for our purpose. Indeed it is well known* that for these asymmetric kernels the bilinear expansion formulae

$$(48) \quad K(r, \xi) = \sum_{n=1}^{\infty} \frac{y_n(r) z_n(\xi)}{\alpha_n^2 - \alpha^2}$$

$$(49) \quad G(r, \xi) = \sum_{n=1}^{\infty} \frac{y_n(\xi) z_n(r)}{\alpha_n^2 - \alpha^2}$$

are valid, where the second members are uniformly convergent in the interval $0 - R$ under consideration.

In view of (48) and of the significance of the function $\sigma(r, \lambda)$ from (23) our solution (41) becomes:

$$(50) \quad y(r, \lambda) = - \sum_{n=1}^{\infty} \frac{y_n(r)}{\alpha_n^2 - \alpha^2} \frac{A^2}{\alpha^2} \int_0^R F''(\xi) z_n(\xi) d\xi \\ + \sum_{n=1}^{\infty} \frac{y_n(r)}{\alpha_n^2 - \alpha^2} \int_0^R \xi \psi(\xi, \lambda) z_n(\xi) d\xi.$$

If the integral in the first term of (50) is twice integrated by parts we get:

$$(51) \quad \int_0^R F''(\xi) z_n(\xi) d\xi = \{z_n(R) F'(R) - z'_n(R) F(R)\} + \int_0^R z_n''(\rho) F(\rho) d\rho.$$

Consider the contribution to $y(r, \lambda)$ as given in (50) arising from the expression in brackets in (51). With the aid of (48) it is clear that this contribution may be written in the form:

$$(52) \quad \frac{A^2}{\alpha^2} F'(R) \left\{ K(r, \xi) \right\}_{\xi=R} - \frac{A^2}{\alpha^2} F(R) \left\{ \frac{\partial}{\partial \xi} K(r, \xi) \right\}_{\xi=R}$$

* A. Kneser, *Integralgleichungen*, Section 6.

If we designate by y_1 and y_2 the analytical expressions of the Green function $K(r, \xi)$ in the intervals $(0 - \xi)$ and $(\xi - R)$ we may write:

$$(53) \quad \begin{aligned} y_1 &= -C_1 r + M_1 \sin \alpha r + N_1 \cos \alpha r \\ y_2 &= -C_2 r + M_2 \sin \alpha r + N_2 \cos \alpha r. \end{aligned}$$

The six coefficients in y_1 and y_2 may be computed from the following conditions which y_1 and y_2 must evidently satisfy

$$(54) \quad \begin{aligned} y_1(0) &= 0; & y_2(R) &= 0; & y_1(\xi) &= y_2(\xi) \\ y'_1(r, \xi) - y'_2(r, \xi) &= 1 & \text{for } r &= \xi. \end{aligned}$$

$$\int_0^R \rho y_1(\rho, \xi) d\rho = \frac{A^2}{B^2} C_1; \quad \int_0^R \rho y_2(\rho, \xi) d\xi = \frac{A^2}{B^2} C_2.$$

It may be easily verified that for $\xi = R$ the above system of equations yields $C_1 = M_1 = N_1 = 0$. Accordingly the first term in (52) vanishes identically. The vanishing of the second term may be proven in a similar manner. Thus (51) reduces to its second term and therefore

$$(51') \quad \int_0^R F''(\xi) z_n(\xi) d\xi = \int_0^R z_n''(\xi) F(\xi) d\xi = -\alpha_n^2 \int_0^R F(\xi) z_n(\xi) d\xi.$$

With the aid of (51') and (23) our solution (50) becomes after an obvious transformation:

$$(55) \quad \begin{aligned} y(r, \lambda) &= -\frac{1}{\lambda} \sum_{n=1}^{\infty} y_n(r) \int_0^R F(\rho) z_n(\rho) d\rho + \sum_{n=1}^{\infty} \frac{y_n(r)}{(\alpha_n/A)^2 + \lambda} \int_0^R F(\rho) z_n(\rho) d\rho \\ &+ \frac{1}{kA^2} \sum_{n=1}^{\infty} \frac{y_n(r)}{(\alpha_n/A)^2 + \lambda} \int_0^R \xi \psi(\xi, \lambda) z_n(\xi) d\xi. \end{aligned}$$

The complete solution of our problem is now obtained by subjecting (55) to the inverse Laplace transformation L^{-1} . Since we evidently have $L^{-1}\{\lambda\} = 1$ the contribution to $T(r, t)$ arising from the first term in (55) reduces to $-F(r)$, provided we assume that the function $F(r)$ is piecewise continuous and twice differentiable, in which case we have:

$$F(r) = \sum_{n=1}^{\infty} y_n(r) \int_0^R F(\rho) z_n(\rho) d\rho.$$

Further since $L^{-1}\left\{\frac{1}{(\alpha_n/A)^2 + \lambda}\right\} = e^{-(\alpha_n/A)^2 t}$ and $L^{-1}\{\psi(r, \lambda)\} = \phi(r, t)$

we have by virtue of Borel's theorem *

* See A. N. L.

$$(56) \quad L^{-1} \left\{ \frac{1}{(\alpha_n/A)^2 + \lambda} \psi(r, \lambda) \right\} = e^{-(\alpha_n/A)^2 t} \int_0^t \phi(r, \eta) e^{(\alpha_n/A)^2 \eta} d\eta.$$

The inversion of (55) thus finally yields:

$$T(r, t) = \frac{1}{r} \sum_{n=1}^{\infty} e^{-(\alpha_n/A)^2 t} y_n(r) \left[\int_0^R F(\rho) z_n(\rho) d\rho + \frac{1}{kA^2} \int_0^R \rho z_n(\rho) \left\{ \int_0^t \phi(\rho, \eta) e^{(\alpha_n/A)^2 \eta} d\eta \right\} d\rho \right]$$

substituting for $y_n(r)$ and $z_n(r)$ the values from (25') and (34') and computing the value of the normalizing factor from (39) the last equation becomes after a slight transformation:

$$\begin{aligned} T(r, t) = & \frac{2}{r} \sum_{n=1}^{\infty} \frac{e^{-(\alpha_n/A)^2 t} \left(\sin \alpha_n r - \frac{r}{R} \sin R \alpha_n \right)}{R - 2 \frac{\sin^2 R \alpha_n}{R \alpha_n^2} + \frac{1}{2 \alpha_n} \sin 2R \alpha_n} \int_0^R \rho f(\rho) \sin \rho \alpha_n d\rho \\ & + \frac{2}{A^2 k r} \sum_{n=1}^{\infty} \frac{e^{-(\alpha_n/A)^2 t} \left(\sin r \alpha_n - \frac{r}{R} \sin R \alpha_n \right)}{R - 2 \frac{\sin^2 R \alpha_n}{R \alpha_n^2} + \frac{1}{2 \alpha_n} \sin 2R \alpha_n} \\ & \times \int_0^R \rho \sin \rho \alpha_n \left\{ \int_0^t \phi(\rho, \eta) e^{(\alpha_n/A)^2 \eta} d\eta \right\} d\rho. \end{aligned}$$

In (57) and (8) we have the complete solution of our problem.

It is readily ascertained that for $C_p = C_v$ (57) reduces to the solution given in A. N. L. Furthermore for $\phi(r, t) = 0$, (57) yields at once the solution given by Duhamel.

INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

THE GEOMETRY OF THE WEDDLE MANIFOLD W_p .

By ARTHUR B. COBLE and JOSEPHINE H. CHANLER.

Introduction. The Weddle manifold W_p has been defined¹ to be that manifold of p dimensions ($p \leq 2$) in an odd space S_{2p-1} which is the locus of fixed points of a certain Cremona involution I attached to a symmetric set of $2p+2$ F -points, P_{2p+2}^{2p-1} . The unique rational norm-curve, N_{2p-1}^{2p-1} , on P_{2p+2}^{2p-1} serves as a convenient reference curve for points of the space S_{2p-1} . The importance of W_p is due not merely to its intrinsic geometric interest but also to the fact that W_p is birationally related to the generalized Kummer manifold K_p of Klein and Wirtinger which is defined by the theta squares provided the theta functions of genus p are of *hyperelliptic* type. The primary purpose of this memoir is to study the geometry of W_p itself, but the matters chosen for study are sometimes such as are fundamentally related to the mapping of W_p upon K_p .

It has been shown¹ that the coördinates of a point on W_p can be expressed by means of hyperelliptic theta functions, and that then the theta squares determine on W_p the sections of W_p by the members of a certain mapping system Σ of order p with $(p-1)$ -fold points at P_{2p+2}^{2p-1} . This system Σ maps W_p upon K_p provided the dimension of Σ is $2^p - 1$, the dimension of the space of K_p . In § 1 the complete base of Σ is obtained, the dimension $2^p - 1$ of Σ is verified, and the dimensions of the subsystems of Σ which contain F -spaces of W_p of various kinds are found. When $p > 2$ these subsystems yield "singular spaces" of K_p of novel type.

Algebraic parametric representations of the generic point on W_p are given in § 2, and these are extended in §§ 3, 4, . . . to study certain systems of curves on W_p . A sketch of the content appears in (⁵).

1. The mapping system Σ . We recall [cf. ¹, § 3] the *finite Cremona group*, $G_{2^{2p+1}}$, attached to the figure P_{2p+2}^{2p-1} of $2p+2$ points in S_{2p-1} , say p_1, \dots, p_{2p+2} . If p_3, \dots, p_{2p+2} are taken as reference points, p_1, p_2 as points y, z , the equations of the element I_{12} of $G_{2^{2p+1}}$ are $x_i x'_i = y_i z_i$ ($i = 1, \dots, 2p$). The abelian $G_{2^{2p+1}}$ is generated by elements I_{ij} of this type. We recall also the definition of the F -loci of the elements of this group—in particular the k -th F -locus of the j -th kind, $\pi_{i_1 i_2 \dots i_{2k+2-j}}^{(j)}$ ($j = 1, \dots, p$). This is, when $j \leq k \leq (2p+j)/2$, the locus of dimension $2p-1-j$ which is described by the ∞^{k-j} S_{2p-k-1} 's on $p_{i_{2k+3-j}}, \dots, p_{i_{2p+2}}$ and on $k-j$ variable points of

N^{2p-1} , the norm-curve on P_{2p+2}^{2p-1} ; and the locus has the order $\binom{k}{j}$, the multiplicity $\binom{k}{j}$ on the $S_{2p-2k+j-1}$ defined by $p_{i_{2k+2-j}}, \dots, p_{i_{2p+2}}$ and the multiplicity $\binom{k-1}{j}$ along N^{2p-1} . On the other hand, when $(j-2)/2 \geq k < j$, it is the locus of dimension $j-1$ which is described by the $\infty^{j-k-1} S_k$'s on $p_{i_1}, \dots, p_{i_{2k+2-j}}$ and on $j-k-1$ variable points of N^{2p-1} ; and this locus has the order $\binom{2p-k-1}{j-k-1}$, the multiplicity $\binom{2p-k-1}{j-k-1}$ on the S_{2k-j+1} defined by $p_{i_1}, \dots, p_{i_{2k+2-j}}$ and the multiplicity $\binom{2p-k-2}{j-k-2}$ along N^{2p-1} .

All the F -loci of the j -th kind are conjugate under $G_{2^{2p+1}}$. When $j = p$, they all are of the same dimension $p-1$. When however $j < p$, those for which $k \geq j$ have the dimension $2p-1-j$, and those for which $k < j$ have the smaller dimension $j-1$. With respect to these F -spaces of smaller dimension we state the theorem:

(1) *The mapping system Σ , the system of spreads of order p with $(p-1)$ -fold points at P_{2p+2}^{2p-1} , contains every F -locus for which $k < j < p$ as a basic locus of multiplicity $p-j$. No other points of S_{2p-1} are base points of Σ .*

The first case of this theorem, $k=0, j=1$ restates the defining property of Σ , i. e., that Σ has $(p-1)$ -fold points at P_{2p+2}^{2p-1} . The second case, $k=0, j=2$, states that Σ contains N^{2p-1} as a basic locus of multiplicity $p-2$. We first prove the theorem for this case. Let $(\alpha x)^p = 0$ be a generic member of Σ . We observe then that

(1.1) $(\alpha x)^p = 0$ contains N^{2p-1} if $p > 2$.

For, α cuts N^{2p-1} at P_{2p+2}^{2p-1} in $(2p+2)(p-1) = (2p-1)p + (p-2)$ points, whence α contains N^{2p-1} if $p-2 \geq 1$.

We now prove the lemma:

(1.2) *If q_1, \dots, q_t are any t points ($t \geq p-3$) of P_{2p+2}^{2p-1} , then $\phi_t(x) \equiv (\alpha q_1)(\alpha q_2) \dots (\alpha q_t)(\alpha x)^{p-t} = 0$ has $(p-t)$ -fold points at q_1, \dots, q_t , $(p-t-1)$ -fold points at the remaining points of P_{2p+2}^{2p-1} , and contains N^{2p-1} .*

For, the lemma is true for $t=1$, since $\phi_1(x) = (\alpha q_1)(\alpha x)^{p-1} = 0$ has a $(p-1)$ -fold point at q_1 with the same tangent cone as $(\alpha x)^p = 0$ (thus, according to (1.1), containing the tangent to N^{2p-1} at q_1) and has $(p-2)$ -fold points at the remaining points of P_{2p+2}^{2p-1} . Thus $\phi_1(x)$ contains N^{2p-1} if $p \geq 4$ since it meets N^{2p-1} at P_{2p+2}^{2p-1} in $(2p+1)(p-2) + p = (p-1)(2p-1) + (p-3)$ points. Since the lemma is true for $t=1$, let us assume that it is true for values of t up to $t-1$, i. e., that $\phi_{t-1}(x) \equiv (\alpha q_1) \dots (\alpha q_{t-1})(\alpha x)^{p-t+1}$

$= 0$ has $(p-t+1)$ -fold points at q_1, \dots, q_{t-1} , has $(p-t)$ -fold points at q_t, \dots, q_{2p+2} , and contains N^{2p-1} . Then $\phi_t(x) = (\alpha q_1) \cdots (\alpha q_t)(\alpha x)^{p-t} = 0$ has $(p-t)$ -fold points at q_1, \dots, q_t , $(p-t-1)$ -fold points at q_{t+1}, \dots, q_{2p+2} , and has at q_t the same tangent cone as $\phi_{t-1}(x)$. Hence $\phi_t(x)$ touches N^{2p-1} at q_t , and, by virtue of its symmetry, at q_1, \dots, q_{t-1} also. Thus $\phi_t(x) = 0$ meets N^{2p-1} at P_{2p+2}^{2p-1} in $t(p-t) + t + (2p+2-t)(p-t-1) = (p-t)(2p-1) + (p-t-2)$ points. Hence $\phi_t(x)$ contains N^{2p-1} if $p-t-2 \geq 1$, or if $t \leq p-3$.

The proof of (1.2) being thus complete, we observe (for $t = p-3$) that $(\alpha q_1) \cdots (\alpha q_{p-3})(\alpha x)^3 = 0$ contains N^{2p-1} , whence

(1.3) *The third polar of any point on N^{2p-1} as to any member of Σ is apolar to any $p-3$ points of P_{2p+2}^{2p-1} .*

To complete the proof of (1) for the case $k=0, j=2$, i. e., that

(1.4) *Every member of Σ contains N^{2p-1} to multiplicity $p-2$ at least,*

we take $2p$ points of the set P_{2p+2}^{2p-1} as reference points. Then $(\alpha x)^p$ has the form $\Sigma a_{i_1 \dots i_p} x_{i_1} \dots x_{i_p}$, since the reference points are $(p-1)$ -fold. If y is any point on N^{2p-1} , the polar $(\alpha y)^3(\alpha x)^{p-3}$ has the form $\Sigma b_{i_1 \dots i_{p-3}} x_{i_1} \dots x_{i_{p-3}} = 0$. But, according to (1.3), every $b_{i_1 \dots i_{p-3}}$ is zero; whence $(\alpha y)^3(\alpha x)^{p-3} \equiv 0$ in x ; i. e. y on N^{2p-1} is a $(p-2)$ -fold point.

The basic loci of Σ , $\pi_{(k)}^{(j)}$, mentioned in (1) are defined by the inequalities,

$$(1.5) \quad (j-2)/2 \leq k < j \leq p-1.$$

All the F -loci of the j -th kind constitute a conjugate set under $G_{2^{2p+1}}$; and in such a set the basic F -loci are distinguished from the others by the fact that their dimension is $j-1$ rather than $2p-j-1$. It is convenient to represent these loci by the notation,

$$(1.6) \quad \pi_{(k)}^{(j)} = S_k(q^{2k+2-j} z^{j-k-1}),$$

which indicates a locus of S_k 's on some selected set of $2k+2-j$ points q in P_{2p+2}^{2p-1} and on $j-k-1$ variable points z on N^{2p-1} .

We now examine the generic point on $\pi_{(k)}^{(j)}$ in (1.6) which can be represented as

$$z = \lambda_1 q_1 + \cdots + \lambda_{2k+2-j} q_{2k+2-j} + \mu_1 z_1 + \cdots + \mu_{j-k-1} z_{j-k-1}.$$

This is a $(p-j)$ -fold point of $(\alpha x)^p = 0$ if $(\alpha z)^{j+1}(\alpha x)^{p-j-1} \equiv 0$. This $(j+1)$ -th polar of z has the form,

$$\Sigma c(\alpha q_1)^{r_1} \cdots (\alpha q_{2k+2-j})^{r_{2k+2-j}} (\alpha z_1)^{s_1} \cdots (\alpha z_{j-k-1})^{s_{j-k-1}} (\alpha x)^{p-j-1},$$

where $r_1 + \cdots + r_{2k+2-j} + s_1 + \cdots + s_{j-k-1} = j+1$. Since $(\alpha q_i)^2 (\alpha x)^{p-2} \equiv 0$ [cf. (1)], and $(\alpha z_i)^3 (\alpha x)^{p-3} \equiv 0$ [cf. (1.4)], a term of this polar vanishes unless in it every $r_i \geq 1$ and every $s_i \geq 2$. For the non-vanishing terms, therefore, $\sum r_i + \sum s_i \geq (2k+2-j) + 2(j-k-1) = j$. Since the sum must be $j+1$, there are no non-vanishing terms. Thus the proof of the multiplicity statement in (1) is complete.

In order to prove that Σ has no other basis points, we observe first that (1.7) *All of the basic F -loci of Σ in (1) are contained in the basic F -loci of the kind $j = p-1$.*

For, if $j' (< j)$ and k' satisfy the inequalities (1.5), and if we take $k' = j' - a$ ($a \geq 1$), then we can take $k = j - a$. The F -locus $S_{k'}(q^{2k'+2-j'}z^{j'-k'-1})$ is then contained in the F -locus $S_k(q^{2k+2-j}z^{j-k-1})$. Indeed $2k' + 2 - j' = j' - 2a + 2 < 2k + 2 - j = j - 2a + 2$, and $j' - k' - 1 = a - 1 = j - k - 1$. Hence each $S_{k'}$ on the one locus is contained in an S_k on the other locus. Thus the basic F -loci of kind j' are all contained on those of kind $j = p-1$ of maximum dimension. We have thus only to prove that every basis point of Σ is on a basic F -locus of kind $j = p-1$.

Included among the F -loci are those of the first kind, $j = 1$. These have a somewhat exceptional position. For $k = 0$, they are basic, being the sets of directions about each of the points of P^{2p-1}_{2p+2} . For larger values of $k \geq p$ they are the P -loci, or *principal manifolds*, of the elements of the Cremona $G_{2^{2p+1}}$. They are paired in such wise that the members of a pair make up one of 2^{2p} degenerate members of the mapping system Σ [cf. ¹, § 5, (31a)]. We have listed these pairs in the table (1.8) below. Opposite them are listed the basic F -loci for $j = p-1$. We prove that Σ has no other basis points by showing that the F -loci listed are the only points common to all the degenerate members of Σ that are listed.

The table is as follows:

<i>Degenerate members of Σ</i>		<i>Basic loci ($j = p-1$)</i>	
1 :	$S_{p-1}(q^1 z^{p-1})$	1 :	$S_{p-2}(\gamma^{p-1})$
2 :	$S_p(q^3 z^{p-2}) \cdot S_{2p-2}(q'^{2p-1})$	2 :	$S_{p-3}(\gamma^{p-3} z)$
.	.	.	.
$k :$	$S_{p-2+k}(q^{2k-1} z^{p-k})$.	.
(1.8)	$\cdot S_{2p-k}(q'^{2p+3-2k} z^{k-2})$	$l :$	$S_{p-1-l}(\gamma^{p+1-2l} z^{l-1})$
.	.	.	.
.	$(p+1)/2 : S_{(3p-3)/2}(q^p z^{(p-1)/2})$.	.
.	$\cdot S_{(3p-1)/2}(q'^{p+2} z^{(p-3)/2})$	$(p+1)/2 :$	$S_{(p-3)/2}(z^{(p-1)/2})$
.	$(p+2)/2 : S_{(3p-2)/2}(q^{p+1} z^{(p-2)/2})$.	.
.	$\cdot S_{(3p-2)/2}(q'^{p+1} z^{(p-2)/2})$	$p/2 :$	$S_{(p-2)/2}(\gamma z^{(p-2)/2})$

Here we use the first or the second of the last two lines according as p is odd or even. In the first column q^i is any set of i points selected from P_{2p+2}^{2p-1} and q'^{2p+2-i} is the complementary set of $2p+2-i$ points. In the second column the points r are also selected from P_{2p+2}^{2p-1} . In both columns the z 's are variable points on N^{2p-1} .

We observe first that the set 1 of degenerate members of Σ will have in common only the points of $S_{p-2}(z^{p-1})$. For, a point P in S_{2p-1} on $S_{p-1}(p_1 z^{p-1})$ and $S_{p-1}(p_2 z^{p-1})$ is represented by a binary $(2p-1)$ -ic which is expressible in two ways as a sum of p $(2p-1)$ -th powers. But, there being no identity connecting $2p$ distinct $(2p-1)$ -th powers, the coefficients of p_1 and p_2 in the two expressions must vanish, and the points z_1, \dots, z_{p-1} in the two expressions, as well as their coefficients, must coincide, i. e. P is a point on a $(p-1)$ -secant S_{p-2} of N^{2p-1} . We therefore examine for base points only the multi-secant spaces of N^{2p-1} of the dimensions contained in the second column of (1.8) and prove that they are basic only when the number of variable points z is precisely the number indicated.

Consider the degenerate member k of Σ and the basic locus l . This has been shown to be on all of the members of Σ . Consider however the $S_{p-1-l}(r^{p-2l} z^l)$, which arises from l by changing one fixed r to a variable z , with reference to k . If $k-l$ of the points r are found in q , the remaining $p-k-l$ points r and l points z can be found among the $p-k$ variable points z of the first factor of k and $S_{p-1-l}(r^{p-2l} z^l)$ is contained on this first factor. If however only $k-l-1$ of the points r are in q , the S_{p-1-l} is not contained in the first factor. The remaining $p-k-l+1$ points r are already contained among the points q' of the second factor, but the $k-2$ points z of the second factor can not be so disposed as to include the first $k-l-1$ points r and the l variable points z so that $S_{p-1-l}(r^{p-2l} z^l)$ is not contained on certain k 's and is therefore not basic. Hence the basis S_{p-1-l} has at most $l-1$ variable points on N^{2p-1} as in (1.8) and the proof of (1) is complete.

(2) *The dimension of the linear system Σ of spreads of order p with $(p-1)$ -fold points at P_{2p+2}^{2p-1} in S_{2p-1} is $2^p - 1$.*

We take as coördinate system in S_{2p-1} the coefficients of a binary $(2p-1)$ -ic, $(\alpha t)^{2p-1} \equiv (\alpha' t)^{2p-1} \equiv \dots$. Perfect powers such as $(tt_1)^{2p-1}$ then determine the points t_1 on N^{2p-1} ; and in particular $t = t_1, \dots, t_{2p+2}$ determine the points of P_{2p+2}^{2p-1} on N^{2p-1} . We set

$$(2.1) \quad (\omega t)^{2p+2} = (tt_1) \cdot (tt_2) \cdot \dots \cdot (tt_{2p+2}).$$

A spread of order p is determined by a form,

$$(2.2) \quad f(s_1^{2p-1}s_2^{2p-1} \dots s_p^{2p-1}) \equiv (\beta_1 s_1)^{2p-1} (\beta_2 s_2)^{2p-1} \dots (\beta_p s_p)^{2p-1},$$

symmetric in the sets of binary variables s_1, \dots, s_p and of order $2p-1$ in each set. A point represented by the binary $(2p-1)$ -ic above is on this spread if the apolarity condition,

$$(2.3) \quad f(\alpha^{2p-1} \alpha'^{2p-1} \dots \alpha^{(p-1)2p-1}) \equiv (\beta_1 \alpha)^{2p-1} (\beta_2 \alpha')^{2p-1} \dots (\beta_p \alpha^{(p-1)})^{2p-1} = 0,$$

is satisfied. If the binary forms, $(\alpha t)^{2p-1}, (\alpha' t)^{2p-1}, \dots$ are regarded as distinct, the condition (2.3) expresses that the corresponding p distinct points are apolar to the p -ic spread; and in particular the vanishing of (2.2) expresses that the points $t = s_1, \dots, s_p$ on N^{2p-1} are apolar to the p -ic.

We observe first that

(2.4) *A symmetric form, $f(s_1^r s_2^r \dots s_j^r)$, is uniquely determined by its linear covariant, $f(s_1^r s_2^r s_3^r \dots s_j^r)$, to within a symmetric form, $g(s_1^{r-2j+2} \dots s_j^{r-2j+2}) \cdot \Pi(s_m s_n)^2$ [$m < n = 1, \dots, j$], where g is a generic symmetric form of the orders indicated.*

For, if F_1, F_2 are two symmetric forms with this linear covariant, then $(F_1 - F_2)_{s_1=s_2=s} \equiv 0$. Hence $F_1 - F_2$ contains the factor $(s_1 s_2)$. The residual factor must change sign if s_1, s_2 interchange, whence it also contains a factor $(s_1 s_2)$. Thus $F_1 - F_2$, being symmetric, contains the symmetric factor, $\Pi(s_m s_n)^2$, and the residual factor is any symmetric factor of the form g .

(2.5) *A generic symmetric form, $f(s_1^r s_2^r \dots s_j^r)$ contains $\binom{j+r}{r}$ linearly independent coefficients.*

For, it can be interpreted as above as a spread of order j in S_r .

(2.6) *The necessary and sufficient condition that the symmetric form (2.2) represent a member of Σ is that*

$$f(s_1^{2p-1} s_2^{2p-1} s_3^{2p-1} \dots s_p^{2p-1}) \equiv (\omega s)^{2p+2} \cdot (ss_2)^2 \dots (ss_p)^2 \cdot g(s_3^{2p-3} \dots s_p^{2p-3}),$$

where g is a symmetric form of the orders indicated.

For, f in (2.2) being symmetric, it represents a spread of order p , and f in (2.6) represents the second polar of s on N^{2p-1} as to this spread. Since the points P_{2p+2}^{2p-1} are $(p-1)$ -fold on a member of Σ , this polar must vanish identically in s_3, \dots, s_p when $s = t_1, \dots, t_{2p+2}$, and thus the factor $(\omega s)^{2p+2}$ must occur. Conversely,

(a) the symmetry of f in (2.2), and

(b) the occurrence of the factor $(\omega s)^{2p+2}$ in f in (2.6), ensure that f in (2.2) is a member of Σ . Moreover, if in (2.6) we think of s, s_4, \dots, s_p as given, then f is the equation in variable s_3 of the linear polar of s, s, s_4, \dots, s_p on N^{2p-1} . Since N^{2p-1} is a $(p-2)$ -fold curve on Σ , it is a simple curve on the cubic polar of s_4, \dots, s_p ; and the linear polar of s on N^{2p-1} as to this cubic touches N^{2p-1} at s , whence the factor $(s s_3)^2$ occurs.

There still remains the proper determination of g in (2.6) and this determination must arise entirely from (2.6) and the symmetry mentioned in (a) above. However, according to (2.4), this g in (2.6) is independent of a generic term in f in (2.2) of the form

$$(2.7) \quad g(s_1^1 s_2^1 \dots s_p^1) \cdot \Pi(s_m s_n)^2$$

with $p+1$ linearly independent terms. From (a) and (2.6) there follows that

$$\begin{aligned} f(s_1^{2p-1} s_2^{2p-1} r^{2p-1} r^{2p-1} s_5^{2p-1} \dots s_p^{2p-1}) \\ \equiv (\omega r)^{2p+2} \cdot (rs_1)^2 (rs_2)^2 (rs_5)^2 \dots (rs_p)^2 \cdot g(s_1^{2p-3} s_2^{2p-3} s_5^{2p-3} \dots s_p^{2p-3}). \end{aligned}$$

Setting $s_1 = s_2 = s$ in this, and setting $s_3 = s_4 = r$ in (2.6), and comparing the right members, we find that

$$(2.8) \quad g(r^{2p-3} r^{2p-3} s_5^{2p-3} \dots s_p^{2p-3}) \\ = (\omega r)^{2p+2} \cdot (rs_5)^2 \dots (rs_p)^2 \cdot h(s_5^{2p-5} \dots s_p^{2p-5}),$$

where h is a symmetric form of the orders indicated. It is to be observed that

$$\begin{aligned} f(s^{2p-1} s^{2p-1} r^{2p-1} r^{2p-1} s_5^{2p-1} \dots s_p^{2p-1}) \\ = (\omega r)^{2p+2} \cdot (\omega s)^{2p+2} \cdot (rs)^4 (rs_5)^2 \dots (rs_p)^2 (ss_5)^2 \dots (ss_p)^2 \cdot h \end{aligned}$$

makes complete use of the symmetry of f in s_1, \dots, s_4 so far as coincidences are concerned, since if three of the variables coincide, f vanishes identically, N^{2p-1} being a $(p-2)$ -fold curve of the spread f .

Again h in (2.8) is conditioned by (2.6), but, in passing from g to h , there remains according to (2.4) undetermined in g , a generic form,

$$(2.9) \quad g(s_3^3 s_4^3 \dots s_p^3) \cdot \Pi(s_m s_n)^2 \quad [m < n = 3, \dots, p]$$

which may be taken at random with $\binom{p+1}{3}$ linearly independent terms.

An entirely similar argument applied to h , on setting $s_5 = s_6 = u$, yields

$$(2.10) \quad h(u^{2p-5} u^{2p-5} s_7^{2p-5} \dots s_p^{2p-5}) \\ = (\omega u)^{2p+2} \cdot (us_7)^2 \dots (us_p)^2 \cdot i(s_7^{2p-7} \dots s_p^{2p-7}),$$

h being determined by this form i to within a form,

$$(2.11) \quad g(s_5^5, \dots, s_p^5) \cdot \Pi(s_m s_n)^2 \quad [m < n = 5, \dots, p],$$

with $\binom{p+1}{5}$ linearly independent coefficients. On continuing this process we find [cf. (2.7), (2.9), (2.11)] that the requirements (a), (b) determine f in (2.2) only to within $\binom{p+1}{1} + \binom{p+1}{3} + \binom{p+1}{5} + \dots = 2^p$ linearly independent arbitrary coefficients which completes the proof of (2).

This determination of the system Σ suggests the following coordinate system for members of Σ . Let the symmetric form (2.2) which represents a member of Σ be denoted by $f_p^{(2p-1)}$. Let $\Pi_1^p \Omega_{ij}^2$ be the operator which, operating on $f_p^{(2p-1)}$, produces $g_p^{(1)} = (\beta_1 s_1) \cdot \dots \cdot (\beta_p s_p) \cdot \Pi_1^p (\beta_i \beta_j)^2$. We have then a sequence of symmetric forms f , and a symmetric covariant g of each, namely:

$$(2.12) \quad \begin{array}{ll} f_p^{(2p-1)} & : \Pi_1^p \Omega_{ij}^2 (f_p^{(2p-1)}) \equiv g_p^{(1)} \\ [f_p^{(2p-1)}]_{s_p=s_{p-1}=s} / (\omega s)^{2p+2} \cdot \Pi_1^{p-2} (ss_i)^2 \equiv f_{p-2}^{(2p-3)} & : \Pi_1^{p-2} \Omega_{ij}^2 (f_{p-2}^{(2p-3)}) \equiv g_{p-2}^{(3)} \\ [f_{p-2}^{(2p-3)}]_{s_{p-2}=s_{p-3}=s} / (\omega s)^{2p+2} \cdot \Pi_1^{p-4} (ss_i)^2 \equiv f_{p-4}^{(2p-5)} & : \Pi_1^{p-4} \Omega_{ij}^2 (f_{p-4}^{(2p-5)}) \equiv g_{p-4}^{(5)} \\ \cdot & \cdot \\ [f_3^{(p+2)}]_{s_3=s_2=s} / (\omega s)^{2p+2} \cdot (ss_1)^2 \equiv f_1^{(p)} & : f_1^{(p)} \equiv g_1^{(p)} \\ [f_2^{(p+1)}]_{s_2=s_1=s} / (\omega s)^{2p+2} \equiv f_0^{(p-1)} & : f_0^{(p-1)} \equiv g_0^{(p+1)} \end{array}$$

the next to the last, or the last, line being used according as p is odd or even.

Every member of Σ defines uniquely a definite sequence of forms $g_p^{(1)}, g_{p-2}^{(3)}, \dots$, the coefficients of these forms g being 2^p linearly independent combinations of the coefficients of the given member of Σ . Hence

(2.13) *The 2^p arbitrary coefficients of the forms g in (2.12) may be taken as the coordinates of a member of the mapping system Σ in (1).*

If the symmetric form $f_{p-2}^{(2p-3)}$ in (2.12) vanishes identically, all the covariants g except $g_p^{(1)}$ vanish identically and conversely. But if $[f_p^{(2p-1)}]_{s_p=s_{p-1}=s}$ is identically zero for any s , then N^{2p-1} is a $(p-1)$ -fold curve, rather than a $(p-2)$ -fold curve, on the corresponding member of Σ . If the symmetric form $f_{p-4}^{(2p-5)}$ vanishes identically, all the covariants g except $g_p^{(1)}$ and $g_{p-2}^{(3)}$ vanish identically and conversely. Then $[f_p^{(2p-1)}]_{s_p=s_{p-1}=s; s_{p-2}=s_{p-3}=r}$ vanishes identically for every s and r , or the bisecant locus of N^{2p-1} is a $(p-3)$ -fold, rather than a $(p-4)$ -fold, locus of the corresponding member of Σ . In general, then

(3) *The necessary and sufficient condition that a member of the mapping*

system Σ shall belong to the sub-system $\sigma^{(2i)}$ of Σ which contains the F -locus $\pi_{i-1}^{(2i)}$ (the locus of i -secant S_{i-1} 's of $N^{2p-1} = S_{i-1}(z^i)$ [cf. (1.6)]) to the multiplicity $p - 2i + 1$ rather than $p - 2i$ [cf. (1)] is $f_{p-2i}^{(2p-1-2i)} \equiv 0$, or $g_{p-2i'}^{(2i'+1)} \equiv 0$ ($p/2 \geq i' \geq i \geq 1$). The system $\sigma^{(2i)}$ has the dimension $\binom{p+1}{1} + \binom{p+1}{3} + \cdots + \binom{p+1}{2i-1} - 1$. The degenerate members of Σ listed in the table (1.8) which belong to $\sigma^{(2i)}$ are the sets $1, 2, \cdots, i$.

The F -loci of the j -th kind, $j = 2i < p$ are 2^{2p+1} in number, conjugate under $G_{2^{2p+1}}$. However they divide into 2^{2p} pairs, the members of a pair being conjugate under the symmetric element, $I = I_{1,2, \dots, 2p+2}$, in $G_{2^{2p+1}}$. Thus $\pi^{(2i)}$ is paired with $\pi_{1,2, \dots, 2p+2}^{(2i)}$, the locus of $(p-i)$ -secant S_{p-i-1} 's of N^{2p-1} , or $S_{p-i-1}(z^{p-i})$. We now prove that if a member of Σ contains $\pi^{(2i)}$ to multiplicity $p - 2i + 1$ as above, then this member must contain the paired F -locus $\pi_{1,2, \dots, 2p+2}^{(2i)}$ simply, and conversely. For, a point of $\pi^{(2i)}$ is $\lambda_1 r_1 + \cdots + \lambda_i r_i$, and a point of $\pi_{1,2, \dots, 2p+2}^{(2i)}$ is $\mu_1 s_1 + \cdots + \mu_{p-i} s_{p-i}$, the r 's and s 's being generic points of N^{2p-1} . Since N^{2p-1} is a $(p-2)$ -fold curve on $(\alpha x)^p = 0$, a member of Σ , we have the following identities in x, r, s :

$$(a) \quad (\alpha r_j)^3 (\alpha x)^{p-3} \equiv 0, \quad (\alpha s_k)^3 (\alpha x)^{p-3} \equiv 0.$$

If $\pi^{(2i)}$ has multiplicity $p - 2i + 1$ on $(\alpha x)^p = 0$, we have the identity in x, λ, r :

$$(b) \quad (\alpha, \lambda_1 r_1 + \cdots + \lambda_i r_i)^{2i} (\alpha x)^{p-2i} \equiv 0.$$

Now $\pi_{1,2, \dots, 2p+2}^{(2i)}$ is contained simply on $(\alpha x)^p = 0$, if

$$(c) \quad (\alpha, \mu_1 s_1 + \cdots + \mu_{p-i} s_{p-i})^p \equiv 0$$

in s and μ . The identity (c) is satisfied in μ if the identity in the s_1, \dots, s_{p-i} ,

$$(d) \quad (\alpha s_1)^{k_1} (\alpha s_2)^{k_2} \cdots (\alpha s_{p-i})^{k_{p-i}} \equiv 0 \quad (k_1 + \cdots + k_{p-i} = p),$$

is satisfied. According to (a) and (b) we need consider only such terms in (d) as have exponents which satisfy

$$(e) \quad k \geq 2, \quad k_1 + k_2 + \cdots + k_i \geq 2i - 1$$

for any i of the k 's. Suppose that l, m, n of the $p-i$ exponents k have values 0, 1, 2 respectively. Then $l + m + n = p - i$ and $m + 2n = p$ [cf. (d)], whence $n - l = i$. Thus at least i of the k 's have the value 2 and (e) cannot be satisfied (the deficiency on the right being one). Therefore (c) is satisfied by virtue of (a) and (b), and $\pi_{1,2, \dots, 2p+2}^{(2i)}$ is at least a simple locus on

$(\alpha x)^p = 0$. Moreover $\pi_{1, \dots, 2p+2}^{(2i)}$ cannot have a higher multiplicity. Otherwise we should have $l + m + n = p - i$; $m + 2n < p$, $n - l < i$. Thus at most $i - 1$ of the k 's must have the value 2 and (e) can be satisfied. Thus certain terms (d) do not necessarily vanish due to (a) and (b).

Conversely, let $(\alpha x)^p = 0$ contain $\pi_{1, \dots, 2p+2}^{(2i)}$ simply. Then (c) and (d) are satisfied and (a) is satisfied as before. We have then to prove that (b) is satisfied, i. e. that

$$(f) \quad (\alpha r_1)^{l_1} \cdots (\alpha r_i)^{l_i} (\alpha x)^{p-2i} \equiv 0 \quad (l \geq 2; l_1 + \cdots + l_i = 2i),$$

or that

$$(g) \quad (\alpha r_1)^2 \cdots (\alpha r_i)^2 (\alpha x)^{p-2i} \equiv 0.$$

Since $i \geq p/2$, we may in (d) let some of the s 's coincide, if necessary, to get each r twice and thus would have terms in (d) of the form $(\alpha r_1)^2 \cdots (\alpha r_i)^2 \times (\alpha s_1) \cdots (\alpha s_{p-2i}) = 0$. Since this would be valid for all choices of s_1, \dots, s_{p-2i} on N^{2p-1} , it would yield (g).

Since all the 2^p pairs of F -spaces of type $\pi_{1, \dots, 2k}^{(2i)}, \pi_{2k+1, \dots, 2p+2}^{(2i)}$ are conjugate under $G_{2^{p+1}}$, we have proved that

(4) *The system Σ contains 2^{2p} linear sub-systems of type $\sigma_{1, \dots, 2k}^{(2i)} = \sigma_{2k+1, \dots, 2p+2}^{(2i)} \equiv \sigma_{(1, \dots, 2k; 2k+1, \dots, 2p+2)}^{(2i)}$, conjugate under $G_{2^{p+1}}$ and of dimension given in (3). The linear system $\sigma_{(1, \dots, 2k; 2k+1, \dots, 2p+2)}^{(2i)}$ is that sub-system of Σ which contains the pair of F -loci,*

$$\pi_{(1, \dots, 2k; 2k+1, \dots, 2p+2)}^{(2i)} \equiv \pi_{1, \dots, 2k}^{(2i)}, \pi_{2k+1, \dots, 2p+2}^{(2i)},$$

simply, i. e. to a multiplicity one greater than the normal multiplicity for all members of Σ .

We have thus far considered only those sub-systems of Σ which contain, to multiplicity one greater than the normal, the F -loci of even rank $j = 2i$. These have a greater degree of simplicity due to the fact that for each rank $j = 2i$ there is one system, $\sigma_{1, 2, \dots, 2p+2}^{(2i)}$, which is symmetrically related to N^{2p-1} , the corresponding F -loci being loci of multi-secant spaces of N^{2p-1} with no fixed intersections. We now consider the F -loci of odd rank, $F^{(j)}$, $j = 2i - 1$ ($1 \leq i \leq (p+1)/2$). As an example of such an $F^{(j)}$ we take $\pi_1^{(2i-1)} = S_{i-1}(p_1 z^{i-1})$ which is paired with $\pi_{2, \dots, 2p+2}^{(2i-1)} = S_{p-i}(p_1 z^{p-i})$. The locus $\pi_1^{(2i-1)}$ is a basic locus of Σ of multiplicity $p - 2i + 1$ [cf. (1)] except in the end case, (p odd), $i = (p+1)/2$. In this end case the paired loci $\pi_1^{(p)}, \pi_{2, \dots, 2p+2}^{(p)}$ coincide.

We seek as before the dimension of the linear system $\sigma_1^{(2i-1)}$, contained in Σ , which has $\pi_1^{(2i-1)}$ as a locus of multiplicity $p - 2i + 2$, one greater than the normal multiplicity of $\pi_1^{(2i-1)}$ on members of Σ . This will be accomplished by discussing the polar system Σ_1 of p_1 with respect to Σ . We give certain preliminary theorems and lemmas which refer to this polarized system Σ_1 . A first theorem relating to Σ is

(5) *The linear system Σ of p -ics with $(p-1)$ -fold points at P_{2p+2}^{2p-1} has a single member with a p -fold point at p_1 .*

The theorem is obvious when $p=2$ and we assume it true for $p=3, \dots, p-1$. Let M be any member of Σ with a p -fold point at p_1 . Carry out on M the involution $I_{2p+1, 2p+2}$ of $G_{2^{2p+1}}$. Since in general a member of Σ is carried by this I into a member of Σ , this member M with an extra multiplicity at p_1 is transformed into a member M' of Σ which consists of $S_{2p-2}(2, \dots, 2p)$ and a spread $M_{2p-2}[1^{p-1}2^{p-2} \dots (2p)^{p-2}(2p+1)^{p-1}(2p+2)^{p-1}]^{p-1}$. This M_{2p-2} of order $p-1$ with $(p-1)$ -fold points at p_{2p+1}, p_{2p+2} , is a cone with a $(p-1)$ -fold line on these two points. It is therefore the dilation from S_{2p-3} of an M_{2p-4} of order $p-1$ with a $(p-1)$ -fold point at q_1 , and $(p-2)$ -fold points at q_2, \dots, q_{2p} . The theorem being true for $p-1$, this M_{2p-4} is unique, whence M_{2p-2} , and M' , and therefore M , are unique.

As an immediate consequence of (5) and (2), we have

(5.1) *The dimension of the linear system Σ_1 , the polar of p_1 as to Σ , is $2^p - 2$.*

For, in polarizing, only those members of Σ with a p -fold point at p_1 are lost.

(5.2) *If $(\alpha x)^p = 0$ has $(p-2)$ -fold points along a norm-curve N^{2p-1} , and has a $(p-1)$ -fold point at p_1 on N^{2p-1} , then the polar $(\alpha p_1)(\alpha x)^{p-1} = 0$ is a cone of order $p-1$ which contains the tangent to N^{2p-1} at p_1 as a line of $(p-2)$ -fold points.*

For, if $(\alpha x)^p = 0$ be written as in (2.2) the condition that it have N^{2p-1} as a $(p-2)$ -fold curve yields the identity,

$$(a) \quad (\beta_1 s)^{2p-1} (\beta_2 s)^{2p-1} (\beta_3 s)^{2p-1} (\beta_4 s_4)^{2p-1} \dots (\beta_p s_p)^{2p-1} \equiv 0,$$

in s, s_4, \dots, s_p . The condition that it have a $(p-1)$ -fold point at p_1 with parameter $s = t_1$, yields the identity,

$$(b) \quad (\beta t_1)^{2p-1} (\beta_2 t_1)^{2p-1} (\beta_3 s_3)^{2p-1} \dots (\beta_p s_p)^{2p-1} \equiv 0,$$

in s_3, \dots, s_p . Any point along the tangent to N^{2p-1} at t_1 is given by variable r in the $(2p-1)$ -ic, $(t_1 t)^{2p-2} \cdot (rt)$. This point will be a $(p-2)$ -fold point of the polar of p_1 if

$$(c) \quad (\beta_1 t_1)^{2p-1} (\beta_2 t_1)^{2p-2} (\beta_2 r) (\beta_3 t_1)^{2p-2} (\beta_3 r) (\beta_4 s_4)^{2p-1} \dots (\beta_p s_p)^{2p-1} \equiv 0$$

in r, s_4, \dots, s_p . We wish to prove that (c) is a consequence of (a) and (b). Since all three of these contain $(\beta_4 s_4)^{2p-1} \dots (\beta_p s_p)^{2p-1}$, we shall omit these factors in the sequel. The identity (a) expresses for arbitrary s_4, \dots, s_p that the form in s vanishes identically. It vanishes therefore for every k and r in $s = t_1 + kr$. On making this substitution in (a), and taking account of the symmetry in the β 's, the coefficient of k^2 yields

$$3 \binom{2p-1}{2} (\beta_1 t_1)^{2p-1} (\beta_2 t_1)^{2p-1} (\beta_3 t_1)^{2p-3} (\beta_3 r)^2 \\ + 3 \binom{2p-1}{1}^2 (\beta_1 t_1)^{2p-1} (\beta_2 t_1)^{2p-2} (\beta_2 r) (\beta_3 t_1)^{2p-2} (\beta_3 r) \equiv 0.$$

The first term of this vanishes due to (b) on replacing in (b) the arbitrary s_3^{2p-1} by $t_1^{2p-3} r^2$; the second term is (c).

Another necessary lemma is

(5.3) *If M is a member of Σ with only a $(p-1)$ -fold point at p_1 [cf. (5)], and if M_1 is the polar of p_1 with respect to M , then any linear S_r on p_1 which is k -fold on M is k -fold on M_1 ; conversely if an S_{r-1} is k -fold on M_1 and on M , then the $S_r = [S_{r-1}, p_1]$ is k -fold on M .*

It is sufficient to prove this for an $S_1 = yp_1$ on p_1 . Let M, M_1 be $(\alpha x)^p = 0$, $(\alpha p_1)(\alpha x)^{p-1} = 0$. Since p_1 is a $(p-1)$ -fold point of M , (a) $(\alpha p_1)^2(\alpha x)^{p-2} \equiv 0$. The S_1 is k -fold on M if (b) $(\alpha, y + \lambda p_1)^{p-k+1}(\alpha x)^{k-1} \equiv 0$; or, making use of (a), if (c) $(\alpha y)^{p-k+1}(\alpha x)^{k-1} + (p-k+1)\lambda(\alpha p_1)(\alpha y)^{p-k}(\alpha x)^{k-1} \equiv 0$. This being true for any λ , $(\alpha p_1)(\alpha y)^{p-k}(\alpha x)^{k-1} \equiv 0$, i. e., the polar $(\alpha p_1)(\alpha x)^{p-1}$ has k -fold points at points y on yp_1 . Conversely, if y is k -fold on M and M_1 , then each term of (c) vanishes, and (b) is satisfied for every point on yp_1 .

We now consider the system in Σ with only a $(p-1)$ -fold point at p_1 . Its dimension is $2p-2$ and it contains the basic F -locus $\pi_1^{(2i-1)}$ to multiplicity $p-2i+1$. The polar system Σ_1 has the order $p-1$ and the following multiplicities: $p-1$ at p_1 ; $p-2$ at p_2, \dots, p_{2p+2} , along lines $p_1 p_i$ [cf. (5.3)], and along the tangent to N^{2p-1} at p_1 [cf. (5.2)]; $p-3$ along N^{2p-1} and on $\pi_1^{(3)} = S_1(p_1 z)$; and $p-2i+1$ along the basic F -locus $\pi_1^{(2i-1)} = S_{i-1}(p_1 z^{i-1})$. Since the system Σ_1 has order $p-1$ and multiplicity $p-1$ at p_1 , it is a system of cones defined completely by p_1 and by its section Σ'_1 by an S'_{2p-2} not on p . We examine this system Σ'_1 . It has the order $p-1$, the multi-

plicity $p-3$ along the N^{2p-2} which is the projection of N^{2p-1} from p_1 upon S'_{2p-2} , and the multiplicity $p-2$ at the set of points Q_{2p+2}^{2p-2} on N^{2p-2} which is the projection from p_1 of the set P_{2p+2}^{2p-1} on N^{2p-1} . We now prove as for (2.6) that

(5.4) *The necessary and sufficient condition that a symmetric form represent a member of Σ'_1 is that*

$$f(s^{2p-2}s^{2p-2}s_3^{2p-2}\cdots s_{p-1}^{2p-2}) \equiv (\omega s)^{2p+2} \cdot (ss_3)^2 \cdots (ss_{p-1})^2 \cdot f_1(s_3^{2p-4}, \cdots, s_{p-1}^{2p-4})$$

where f_1 is a symmetric form of the orders indicated.

This condition utilizes explicitly only that N^{2p-2} is $(p-3)$ -fold, and that the points Q_{2p+2}^{2p-2} determined by $(\omega s)^{2p+2} = 0$ are $(p-2)$ -fold on Σ'_1 . The occurrence of the factors $(ss_3)^2, \cdots, (ss_{p-1})^2$ follows as before. In passing from the symmetric form f as in (5.4) to the symmetric form f_1 there is lost, according to (2.4), a symmetric form

$$(a) \quad (s_1s_2)^2 \cdots (s_{p-2}s_{p-1})^2 \cdot g(s_1^2s_2^2 \cdots s_{p-1}^2)$$

for which the $\binom{p+1}{2}$ coefficients of g may be taken arbitrarily without affecting the defining properties of the member of Σ'_1 or of f_1 . The only conditions on f_1 in (5.4), as in the earlier case (2.8), are those embodied in (5.4), and in the original symmetry, which yield for f_1 the condition,

$$(b) \quad f_1(s^{2p-4}s^{2p-4}s_5^{2p-4}\cdots s_{p-1}^{2p-4}) \\ \equiv (\omega s)^{2p+2} \cdot (ss_5)^2 \cdots (ss_{p-1})^2 \cdot f_2(s_5^{2p-6} \cdots s_{p-1}^{2p-6}),$$

and which leave undetermined in f_1 a symmetric form,

$$(c) \quad (s_3s_4)^2 \cdots (s_{p-2}s_{p-1})^2 \cdot g_1(s_3^4, \cdots, s_{p-1}^4)$$

for which the $\binom{p+1}{4}$ coefficients of g_1 may be taken arbitrarily.

Continuing in this fashion we find that

(5.5) *The dimension of the system Σ'_1 in S'_{2p-2} of order $p-1$ with $(p-2)$ -fold points at Q_{2p+2}^{2p-2} and $(p-3)$ -fold curve N^{2p-2} is $\binom{p+1}{2} + \binom{p+1}{4} + \cdots - 1 = 2^p - 2$.*

On comparing this with (5.1) we see that

(5.6) *The polar system Σ_1 of p_1 as to Σ is the conical dilation into S_{2p-1} with vertex p_1 in S_{2p-1} of the system Σ'_1 in S'_{2p-2} described in (5.5).*

The method of derivation of the dimension of Σ'_1 in (5.5) yields a di-

vision of Σ into sub-systems. First there is the unique member of Σ [cf. (5)] which is lost in the polar system Σ_1 and therefore in the section Σ'_1 . This is the sub-system of Σ of dimension $\binom{p+1}{0} - 1$ which contains the F -locus $\pi_1^{(1)} = S_0(p_1)$ to multiplicity p rather than $p - 1$. This member is defined by $g \equiv 0$; $g_1 \equiv 0, \dots$. In passing from any sub-system of Σ'_1 to a sub-system of Σ , this member must be added. Consider next the sub-system of Σ'_1 defined by the identical vanishing of f_1 in (5.4), or by the identical vanishing of the sequence of forms, $g_1 \equiv 0$, $g_2 \equiv 0, \dots$. If f_1 in (5.4) vanishes identically, the curve N^{2p-2} is $(p-2)$ -fold, rather than $(p-3)$ -fold, on members of Σ'_1 . On applying (5.3), we find that this sub-system of Σ'_1 yields that sub-system of Σ which contains $\pi_1^{(3)} = S_1(p_1 z)$ to multiplicity $p-2$ rather than $p-3$. Its dimension in S'_{2p-2} is $\binom{p+1}{2} - 1$ and thus we find $\binom{p+1}{0} + \binom{p+1}{2} - 1$ as the dimension of the sub-system of Σ which contains $\pi_1^{(3)}$ to multiplicity $(p-3) + 1$.

Continuing in this fashion we have the analog of (3) namely:

(6) *The necessary and sufficient condition that a member of Σ belong to the sub-system, $\sigma_1^{(2i-1)}$, of Σ which contains the basic F -locus $\pi_1^{(2i-1)} = S_{i-1}(p_1 z^{i-1})$ to the multiplicity $p - 2i + 2$ rather than $p - 2i + 1$ is the identical vanishing of the form f_{i-1} [cf. (5.4) (b)], or of the sequence of forms, $g_{i-1}, g_i, g_{i+1}, \dots$ [cf. (5.4) (a), (c)], these forms being determined by the system Σ'_1 in S'_{2p-2} . The dimension of $\sigma_1^{(2i-1)}$ is $\binom{p+1}{0} + \binom{p+1}{2} + \binom{p+1}{4} + \dots + \binom{p+1}{2i-2} - 1$.*

The F -loci for odd j are also paired into 2^{2p} pairs of type

$$(6.1) \quad \pi_{(1, \dots, 2k+1; 2k+2, \dots, 2p+2)}^{(2i-1)} = \pi_{1, \dots, 2k+1}^{(2i-1)} \pi_{2k+2, \dots, 2p+2}^{(2i-1)}$$

the members of a pair being conjugate under $I = I_1, \dots, I_{2p+2}$, and the 2^{2p} pairs being conjugate under $G_{2^{2p+1}}$. It may be proved by the method preceding (4) that the linear sub-system $\sigma_1^{(2i-1)}$ of Σ which contains the basic locus $\pi_1^{(2i-1)}$ to multiplicity $p - 2i + 2$ rather than to the normal multiplicity $p - 2i + 1$ for Σ also contains *simply* the paired F -locus, $\pi_{2,3, \dots, 2p+2}^{(2i-1)}$, which is not basic for Σ , and conversely. We have then the analog of (4), namely:

(7) *The system Σ contains 2^{2p} linear sub-systems of type,*

$$\sigma_{(1, \dots, 2k+1; 2k+2, \dots, 2p+2)}^{(2i-1)} = \sigma_{1, \dots, 2k+1}^{(2i-1)} = \sigma_{2k+2, \dots, 2p+2}^{(2i-1)}$$

conjugate under $G_{2^{2p+1}}$, and of dimension given in (6). The linear system $\sigma_{(1, \dots, 2k+1; 2k+2, \dots, 2p+2)}^{(2i-1)}$ is that sub-system of Σ which contains the pair of F -loci given in (6.1) SIMPLY, i. e., to a multiplicity one greater than the normal multiplicity of the locus for all members of Σ .

When $j = p$ it appears from the definitions of the F -loci given at the outset that the paired F -loci as defined in (4) and (6.1) coincide. Also these F -loci are not basic for Σ . When $2i = p$ in (3), and $2i - 1 = p$ in (6), the dimensions given both become $2^p - 2$. Thus

(8) *It is a single condition on the members of Σ to contain one of the 2^{2p} F -loci of the p -th kind simply.*

We find in (7) an instance of increased simplicity of statement when the F -loci are brought in paired as in (3) and (6). Another instance is embodied in the theorem:

(9) *A pair of F -loci $\pi^{(j)}$ and a pair of F -loci $\pi^{(j+1)}$ are INCIDENT if the division of indices which determines the one can be converted into the division of indices which determines the other by shifting an index from one of the two sets into the other set. Thus a pair $\pi^{(j)}$ contains $2p + 2$ pairs π^{j+1} ($j < p$), and is contained in $2p + 2$ pairs $\pi^{(j-1)}$ ($j > 1$).*

Because of the conjugacy of the pairs under $G_{2^{2p+1}}$ it will be sufficient to prove this for one pair for given j and because of the symmetry of the F -loci in the pair it will be sufficient to examine one index in either set. For each value of j there are linear F -loci, and we take such a typical case, namely:

$$(a) \quad \pi_{(1, \dots, j+2; j+3, \dots, 2p+2)}^{(j)} = S_{2p-j-1}(p_{j+3} \cdots p_{2p+2}), S_p(p_1 \cdots p_{j+2} z^{p-j-1}).$$

We compare this with

$$(b) \quad \pi_{(1, \dots, j+3; j+4, \dots, 2p+2)}^{(j+1)} = S_{2p-j-2}(p_{j+4} \cdots p_{2p+2}), S_p(p_1 \cdots p_{j+3} z^{p-j-2}),$$

and with

$$(c) \quad \pi_{(1, \dots, j+1; j+2, \dots, 2p+2)}^{(j-1)} = S_{2p-j}(p_{j+2} \cdots p_{2p+2}), S_p(p_1 \cdots p_{j+1} z^{p-j}).$$

The first member of (b) is incident with the first member of (a); the second member of (b) is incident with the second member of (a), one z in (a) being fixed at p_{j+3} . The first member of (c) is incident with the first member of (a); the second member of (c) is incident with the second member of (a). Since the shifting of an index from one set to the other can be done in $2p + 2$ ways, the incidences of the theorem are established.

The results obtained in this section lead to certain conclusions with respect to the *hyperelliptic* Kummer manifold K_p in S_{2^p-1} . Since Σ contains members which represent on W_p the theta squares which define K_p , and since the dimension of Σ is $2^p - 1$, then Σ maps W_p upon K_p . In this mapping the pairs of F -loci of the first kind contribute members of the mapping system

which pass into the 2^{2p} singular spaces of K_p of dimension $2^p - 2$. On the other hand the 2^{2p} F -loci of the p -th kind map into the 2^{2p} singular points of K_p [cf. (8) and ¹, (66)]. These are the only singular spaces arising from the classic theory. There remain the 2^{2p} pairs of F -loci of kind j ($1 < j < p$) of W_p . We find in § 3 that these pairs of F -loci meet W_p in manifolds of dimension $p - j$, a property which carries over to K_p . The sub-systems of Σ on a pair of F -loci of kind j yield systems of linear spaces in S_{2^p-1} which have for base a singular space of K_p of the j -th kind. Thus a translation of the results obtained above is the following:

(10) *The hyperelliptic K_p in S_{2^p-1} has p systems of singular linear spaces $\Sigma^{(j)}$ ($j = 1, \dots, p$), each system having 2^{2p} members. Each space $\Sigma^{(j)}$ has in common with K_p a manifold of dimension $p - j$. The dimension of a linear space $\Sigma^{(j)}$ is $2^p - 2$ less the dimension of the system $\sigma^{(j)}$ as given in (3) and (6). Each space $\Sigma^{(j)}$ contains $2p + 2$ spaces $\Sigma^{(j+1)}$, and is contained in $2p + 2$ spaces $\Sigma^{(j-1)}$. The spaces $\Sigma^{(j)}$ and spaces $\Sigma^{(p-j)}$ are conjugate under the correlation $G_{2,2^{2p}}$ of K_p .*

Thus K_3 in S_7 has 4^3 singular S_6 's, 4^3 singular S_5 's, and 4^3 singular S_4 's; K_4 in S_{15} has 4^4 singular S_{14} 's, 4^4 singular S_{10} 's, 4^4 singular S_8 's, and 4^4 singular S_6 's; etc. These intermediate singular spaces arise from degenerations of loci on the generic K_p . For example, two singular S_6 's of K_3 in S_7 meet K_3^{24} in an elliptic E_6 in their common S_5 [cf. ³, p. 188 (3)]. When K_3 is hyperelliptic, this (for proper choice of the two singular S_6 's) breaks up into two N^3 's with two common points. The two S_5 's containing these N^3 's are singular S_3 's. The two common points of the two N^3 's arise from the extra zero of the two thetas which define the two S_6 's, these extra zeros being characteristic of the hyperelliptic case.

2. Parametric forms of the hyperelliptic Weddle p -way in S_{2p-1} . The hyperelliptic Weddle p -way in S_{2p-1} has been defined [cf. ¹, (34)] as the locus of fixed points of the involution $I = I_1, \dots, I_{2p+2}$ in the $G_{2^{2p+1}}$ determined by the set of points P_{2p+2}^{2p-1} in S_{2p-1} . As x varies on W_p , the set of points, P_{2p+3}^{2p-1} , consisting of P_{2p+2}^{2p-1} and x , has been shown to be "associated" to the set of points R_{2p+3}^2 which consists of the $2p + 2$ branch points and the multiple point O of a planar hyperelliptic curve H_p of order $p + 2$ with p -fold point at O , and with $2p + 2$ branch lines on O whose parameters are projective to the parameters of P_{2p+2}^{2p-1} on their norm-curve N^{2p-1} . We examine this association.

Let the hyperelliptic curve H_p have the equation,

$$(1) \quad H_p = y_0^2 f_p(y_1, y_2) + 2y_0 f_{p+1}(y_1, y_2) + f_{p+2}(y_1, y_2) = 0,$$

where f_p, f_{p+1}, f_{p+2} are binary forms in y_1, y_2 of orders indicated by the subscripts. We set

$$(2) \quad (\omega t)^{2p+2} \equiv f_{p+1}^2(t_1 t_2) - f_p(t_1 t_2) \cdot f_{p+2}(t_1 t_2) \equiv (tt_1)(tt_2) \cdot \dots (tt_{2p+2}), \\ z_t \equiv \sqrt{(\omega t)^{2p+2}}.$$

In terms of this irrationality, z_t , a parametric equation of H_p is

$$(3) \quad y_0 : y_1 : y_2 = -f_{p+1}(t_1, t_2) + z_t : t_1 f_p(t_1, t_2) : t_2 f_p(t_1, t_2),$$

the parameter being $t_1 : t_2 = t$. When $(tt_i) = 0$, we have a branch point of the g_1^2 on H_p with coördinates

$$(4) \quad y_0 : y_1 : y_2 = -f_{p+1}(t_{i1}, t_{i2}) : t_{i1} f_p(t_{i1}, t_{i2}) : t_{i2} f_p(t_{i1}, t_{i2}) \\ [i = 1, \dots, 2p + 2].$$

The p -fold point O of H_p has coördinates

$$(5) \quad y_0 : y_1 : y_2 = 1 : 0 : 0.$$

Thus the $2p + 2$ branch points and O , the set R^2_{2p+3} in S_2 , have as matrix of coördinates (written vertically and with non-homogeneous parameter t) the following:

$$(6) \quad \begin{array}{ccccccc} -f_{p+1}(t_1) & -f_{p+1}(t_2) & -f_{p+1}(t_3) & : & -f_{p+1}(t_{2p+2}) & & 1 \\ t_1 f_p(t_1) & t_2 f_p(t_2) & t_3 f_p(t_3) & : & t_{2p+2} f_p(t_{2p+2}) & & 0 \\ f_p(t_1) & f_p(t_2) & f_p(t_3) & : & f_p(t_{2p+2}) & & 0. \end{array}$$

Using as a coördinate system in S_{2p-1} the coefficients of a $(2p - 1)$ -ic referred to the N^{2p-1} on P^{2p-1}_{2p+2} , then the $2p + 3$ points consisting of P^{2p-1}_{2p+2} and x on W_p have for coördinates:

$$(7) \quad (tt_1)^{2p-1}, (tt_2)^{2p-1}, (tt_3)^{2p-1}, \dots, (tt_{2p+2})^{2p-1}, (\alpha t)^{2p-1}.$$

Here the $(2p - 1)$ -ic, $(\alpha t)^{2p-1}$ is to be determined in such wise that the row product of the row (7), each term with appropriate constant factor, with each of the three rows in (6) is to vanish identically in t , these being the conditions that the two sets of $2p + 3$ points be associated [cf. ³, § 13].

We use the notation,

$$(8) \quad g'_m(s_k) = (s_k s_1)(s_k s_2) \cdot \dots (s_k s_{k-1})(s_k s_{k+1}) \cdot \dots (s_k s_m) \quad [k = 1, \dots, m],$$

in connection with a form $g_m(t) = (ts_1) \cdot \dots (ts_m)$. We also express $f_p(t)$ in factored form as follows:

$$(9) \quad f_p(t) \equiv (tr_1)(tr_2) \cdot \dots (tr_p).$$

The $(2p+2)$ $2p$ -th powers of $(tt_1), \dots, (tt_{2p+2})$ are related by the identity:

$$(10) \quad \sum_{i=1}^{2p+2} (t_i t)^{2p} / \omega'(t_i) \equiv 0.$$

The $(3p+2)$ $3p$ -th powers of $(tt_1), \dots, (tt_{2p+2}), (tr_1), \dots, (tr_p)$ are related by the identity:

$$(11) \quad \sum_{i=1}^{2p+2} (tt_i)^{3p} / f_p(t_i) \cdot \omega'(t_i) + \sum_{h=1}^p (tr_h)^{3p} / \omega(r_h) \cdot f'_p(r_h) \equiv 0.$$

We observe that, due to the identity (10), the row product of (7) and each of the last two rows in (6) is identically zero in t provided that the first $2p+2$ powers in (7) are affected by the following factors respectively:

$$1/f_p(t_1) \cdot \omega'(t_1), \dots, 1/f_p(t_{2p+2}) \cdot \omega'(t_{2p+2}).$$

With these factors we take the row product of (7) and the first row of (6), and find that

$$(12) \quad (\alpha t)^{2p-1} = \sum_{i=1}^{2p+2} f_{p+1}(t_i) \cdot (tt_i)^{2p-1} / f_p(t_i) \cdot \omega'(t_i).$$

If we polarize the identity (11) with respect to $f_{p+1}(t)$, the first sum yields $(\alpha t)^{2p-1}$ in (12). For the second sum we observe that, in the case of the roots r_h of f_p [cf. (2)],

$$(13) \quad z_{r_h} = f_{p+1}(r_h), \quad \omega(r_h) = f_{p+1}^2(r_h).$$

Hence this second sum yields

$$(14) \quad -(\alpha t)^{2p-1} = \sum_{h=1}^p (tr_h)^{2p-1} / z_{r_h} \cdot f'_p(r_h).$$

This formula shows that the coördinates x on W_p are proportional to abelian functions of u_1, \dots, u_p on H_p determined by the p -ad of points on H_p :

$$(15) \quad r_1, z_{r_1}; r_2, z_{r_2}; \dots; r_p, z_{r_p};$$

or by its "superposed" p -ad in which the z 's all change sign. For, the coefficients α in (14) are symmetric in the p pairs of values r_i, z_{r_i} .

Each value of the parameter t determines a pair of points $t, \pm z_t$ [cf. (3)] on H_p . Thus p values of t , say r_1, \dots, r_p , determine p pairs of points on H_p which can be arranged into 2^p p -ads on H_p with one point of a p -ad from each pair. These 2^p p -ads divide into 2^{p-1} pairs of superposed p -ads and determine 2^{p-1} points x on W_p as in (14). The 2^{p-1} points x are obtained in (14) by taking the changes of sign of z_{r_1}, \dots, z_{r_p} . Hence

(16) If x is a point of W_p , the p -secant space S_{p-1} of N^{2p-1} on x meets W_p again in $2^{p-1} - 1$ remaining points which with x form a conjugate set of 2^{p-1} points under the group of order 2^{p-1} in S_{p-1} which consists of the identity and the harmonic perspectivities determined by opposite spaces of the p -edron in S_{p-1} and on N^{2p-1} .

As particular cases for $p = 2, 3, 4$ we may mention:

(17) (a) The bisecant of N^3 on a point $x^{(1)}$ of W_2 in S_3 meets W_2 again in a point $x^{(2)}$ such that $x^{(1)}, x^{(2)}$ are harmonic with the crossings of the bisecant.

(b) The trisecant plane of N^5 on a point $x^{(1)}$ of W_3 in S_5 meets W_3 in four points $x^{(1)}, \dots, x^{(4)}$ whose diagonal triangle is the triad of crossings of the trisecant plane.

(c) The quadri-secant S_3 of N^7 on a point $x^{(1)}$ of W_4 in S_7 meets W_4 in eight points $x^{(1)}, \dots, x^{(8)}$ which make up with the four crossings of S_3 a set of desmic tetrahedra in the S_3 .

We observe also that

(18) The $(2p-1)$ -ic in (14) which represents with respect to N^{2p-1} the point x on W_p defined by H_p in (1) is the $(2p-1)$ -ic which is apolar to f_p and f_{p+1} .

For, the form of the $(2p-1)$ -ic in (14) indicates its apolarity with $f_p = (tr_1) \dots (tr_p)$ [cf. (9)]. If also we operate on the $(2p-1)$ -ic with f_{p+1} , and make use of (13), the result vanishes identically by virtue of the linear relation among the $(p-2)$ -th powers of $(tr_1), \dots, (tr_p)$.

In general this $(2p-1)$ -ic is not also apolar to f_{p+2} . It will be, however, if

$$(19) \quad g_0 f_{p+2} - 2g_1 f_{p+1} + g_2 f_p \equiv 0,$$

where g_0, g_1, g_2 are binary forms in $t_1: t_2$ of the orders indicated. For $p = 2$ this identity can be satisfied for any f_4, f_3, f_2 . For higher values of p it can be satisfied only if the branch points of H_p are on the conic, H_0 :

$$(20) \quad H_0 = g_0 y_0^2 + 2g_1 y_0 + g_2 = 0 \quad [\text{cf. } ^3, \S 38].$$

The two branch points of H_0 are then on H_p . The line joining these two branch points is $g_0 y_0 + g_1 = 0$. This line cuts H_p in p further points. Eliminating y_0 and using (19), the parameters t of the $p+2$ intersections are given by $g_0(g_1^2 - g_0 g_2) f_p \equiv 0$. The p further points are therefore the

further intersections with H_p of the tangents at the p -fold point. Conversely if the p further intersections of these tangents are on a line and we take this line to be $y_0 = 0$, then $f_{p+2} - (at)^2 \cdot f_p \equiv 0$, and the identity (19) is satisfied for the conic $y_0^2 - (at)^2$. The fundamental $(2p+2)$ -ic of branch lines of H_p , and the fundamental quadratic of branch lines of the conic H_0 , are now

$$(21) \quad (\omega t)^{2p+2} = f_{p+1}^2 - (at)^2 \cdot f_p^2 = 0, \quad (at)^2 = 0.$$

In particular for a root t_i of $(\omega t)^{2p+2}$,

$$(22) \quad f_{p+1}(t_i)/f_p(t_i) = \sqrt{(at_i)^2}.$$

We have then, on making use of (12), the theorem:

(23) *If the $2p+2$ branch points of H_p are on a conic, i. e., if the p further intersections of tangents to H_p at the p -fold point are on a line, the point x on W_p determined by H_p is represented by the $(2p-1)$ -ic (with reference to N^{2p-1}),*

$$(\alpha t)^{2p-1} = \sum_{i=1}^{2p+2} (tt_i)^{2p-1} \cdot \sqrt{(at_i)^2/\omega'}(t_i).$$

For variable quadratic, $(at)^2$, this point x runs over a manifold $V_2^{(2p-1)}$ on W_p .

The question naturally arises as to whether the signs of the radicals in (23) may be taken at random if the point x is to remain on $V_2^{(2p-1)}$ on W_p . The following lemma shows that the answer is affirmative:

(24) *Given O and a conic H_0 . Choose any $2p+2$ points s_1, \dots, s_{2p+2} on H_0 , no one the contact of a tangent from O . Let the line pencil from O to s_i have parameters t_i , and let the line t_i cut H_0 in $s_i(+)=s_i$ and $s_i(-)$. Then there exists an H_p with fundamental $(2p+2)$ -ic, t_i , and $2p+2$ branch points s'_i , s'_i being either $s_i(+)$ or $s_i(-)$.*

For, if H_0 is $y_0^2 - y_1 y_2 = 0$, or $y_0 : y_1 : y_2 = s : s^2 : 1$, and if the branch points of H_p are on this conic, then $f_{p+2}(y_1, y_2) = y_1 y_2 f_p(y_1, y_2)$. The H_p can then be written as $(\alpha y)^p \cdot y_0^2 + 2(\beta y)^{p+1} y_0 + y_1 y_2 \cdot (\alpha y)^p = 0$, with $2p+3$ homogeneous parameters in the coefficients of the forms $(\alpha y)^p$, $(\beta y)^{p+1}$ in the binary variables y_1, y_2 . The curve H_p is on the point $s : s^2 : 1$ of H_0 if $(\alpha s^2)^p \cdot s + (\beta s^2)^{p+1} = 0$. The ratios of the coefficients α, β of this equation in s of degree $2p+2$ are uniquely determined by assigning roots s_1, \dots, s_{2p+2} to it, and for each choice of s_i or $-s_i$ we have a curve H_p . On the other hand the fundamental $(2p+2)$ -ic of H_p is $f_{p+1}^2(y_1, y_2) - y_1 y_2 f_p^2(y_1, y_2)$

$= f_{p+1}^2(s^2, 1) - s^2 f_p^2(s^2, 1)$, and it is independent of the choice of s_i or $-s_i$, since $y_1 : y_2 = s^2 : 1$.

If then in (23) the sign of the radical of $(at_i)^2$ be changed, we have a new point of $V_2^{(2p-1)}$ which lies with the original point on a line through p_i on N^{2p-1} . Hence

(25) *For variation of the signs of the radicals in (23) a closed system of $2 \cdot 2^{2p}$ points on $V_2^{(2p-1)}$ is obtained, the system being projected into itself from each of the points of P_{2p+2}^{2p-1} . If the signs of the first two radicals are changed, the point thus obtained is the conjugate of the given point under I_{12} . Thus the closed system consists of two sets of 2^{2p} points conjugate under the Cremona $G_{2^{2p+1}}$ of W_p , depending on the parity of the number of changes of sign.*

Here only the second statement in (25) requires additional proof. The involution I_{12} on W_p corresponds in the plane to the quadratic transformation $A_{012} \cdot (12)$, i. e., to the perspective transformation with center O and F -points, O and the first two branch points [cf. ³, § 38]. Then H_p goes into H'_p and s on H_0 into s on H'_0 . Thus on H'_0 the $2p$ further branch points have parameters s_3, \dots, s_{2p+2} , but the two fixed branch points have parameters $-s_1, -s_2$.

The projection of $V_2^{(2p-1)}$ from p_1 upon the V_2^{2p-2} in S_{2p-2} determined by a set of points P_{2p+1}^{2p-2} upon an N^{2p-2} with parameters t_2, \dots, t_{2p+2} is obtained by taking the linear polar of t_1 as to $(\alpha t)^{2p-1}$ in (23). The resulting $(2p-2)$ -ic with reference to N^{2p-2} determines the projected point in S_{2p-2} . This $(2p-2)$ -ic is

$$(26) \quad \sum_{i=2}^{2p+2} (t_1 t_i) \cdot (t t_i)^{2p-2} \sqrt{(at_i)^2 / \omega'(t_i)}.$$

It has, as is evident, properties entirely analogous to $V_2^{(2p-1)}$ and is, due to the loss of one radical, a doubly covered projection. We have thus confirmed analytically the properties of the manifold $V_2^{(2p-1)}$ on W_p which were obtained in [¹, § 20 and ², § 6] geometrically, the $V_2^{(2p-1)}$ being defined in the first case as the locus of points in S_{2p-1} from which P_{2p+2}^{2p-1} projects into $2p+2$ points in S_{2p-2} on a rational N^{2p-2} , and in the second case as the locus of nodes of degenerate bi-nodal curves in the family of elliptic norm-curves on P_{2p+2}^{2p-1} .

In this case the generalized theorem, due in the case of W_2 to H. F. Baker, applies not to W_p but rather to $V_2^{(2p-1)}$ on W_p .

3. Parametric equations of W_p related to the curves cut out on W_p by $(p+1)$ -secant S_p 's of N^{2p-1} . In the preceding section we have found

that a generic p -secant S_{p-1} of N^{2p-1} cuts W_p in S_{2p-1} in 2^{p-1} points any one of which is generic on W_p , the others forming with this one a symmetric set [cf. 2 (16)]. We now derive certain expressions for the point $(\alpha t)^{2p-1}$ on W_p and also on the section of W_p by a $(p+1)$ -secant S_p of N^{2p-1} . As the dimensions p of the intersecting manifolds indicate, this section is a curve rather than a set of points.

With $(\omega t)^{2p+2} = f_{p+1}^2 - f_p f_{p+2}$, and in terms of the factorizations,

$$(1) \quad \begin{aligned} (\omega t)^{2p+2} &= (tt_1) \cdots (tt_a) \cdots (tt_{2p+2}), \\ f_p(t) &= (tr_1) \cdots (tr_e) \cdots (tr_p), \end{aligned}$$

we have already obtained the following expressions for the point $(\alpha t)^{2p-1}$ on W_p [cf. 2 (12), (13), (14)]:

$$\begin{aligned} (A) \quad (\alpha t)^{2p-1} &= \sum_{d=1}^{d=2p+2} (tt_d)^{2p-1} \cdot f_{p+1}(t_d)/f_p(t_d) \cdot \omega'(t_d); \\ (B) \quad (\alpha t)^{2p-1} &= \sum_{d=1}^{d=2p+2} (tt_d)^{2p-1} \cdot f_{p+2}(t_d)/f_{p+1}(t_d) \cdot \omega'(t_d); \\ (C) \quad -(\alpha t)^{2p-1} &= \sum_{e=1}^{e=p} (tr_e)^{2p-1}/f_{p+1}(r_e) \cdot f'_p(r_e). \end{aligned}$$

The expression (B) is the same as (A), since, for a root t_d of $(\omega t)^{2p+2}$, $f_{p+1}(t_d)/f_p(t_d) = f_{p+2}(t_d)/f_{p+1}(t_d)$.

Suppose now that the line $y_0 = 0$ in the canonical form of H_p^{p+2} is on $j+2$ of the branch points of H_p^{p+2} ($j = -2, -1, 0, \dots, p$). If $j = -2, -1, 0$, this imposes no projective condition on H_p^{p+2} . If however $j = 1, \dots, p$, this requires that H_p^{p+2} be represented on W_p by a point on an F -locus of the j -th kind. Since $y_0 = 0$ cuts H_p^{p+2} in points whose parameters t are given by $f_{p+2}(t) = 0$, the parameters t of these $j+2$ branch points will satisfy both $f_{p+2}(t) = 0$ and $(\omega t)^{2p+2} = 0$, and therefore $f_{p+1}(t) = 0$ also. Let these $j+2$ branch points, say the first $j+2$, be given by $\lambda_{j+2}(t) = 0$. Then we have

$$(2) \quad \begin{aligned} (\omega t)^{2p+2} &= \lambda_{j+2}(t) \cdot \mu_{2p-j}(t), \quad f_{p+2}(t) = \lambda_{j+2}(t) \cdot g_{p-j}(t), \\ f_{p+1}(t) &= \lambda_{j+2}(t) \cdot g_{p-j-1}(t). \end{aligned}$$

In addition to the factorizations (1) we introduce also the following:

$$(3) \quad \begin{aligned} \lambda_{j+2}(t) &= (tt_1) \cdots (tt_a) \cdots (tt_{j+2}), \\ \mu_{2p-j}(t) &= (tt_{j+3}) \cdots (tt_b) \cdots (tt_{2p+2}), \\ g_{p-j-1}(t) &= (ts_1) \cdots (ts_c) \cdots (ts_{p-j-1}). \end{aligned}$$

We remove the factor $\lambda_{j+2}(t)$ from the relation, $(\omega t)^{2p+2} = f_{p+1}^2 - f_p f_{p+2}$, and obtain

$$(4) \quad \pi \equiv \mu_{2p-j} - g_{p-j-1}^2 \cdot \lambda_{j+2} \equiv -f_p \cdot g_{p-j}.$$

If only the roots s_1, \dots, s_{p-j-1} of $g_{p-j-1} = 0$ are given, there still remains an undetermined constant factor in g_{p-j-1} , and thus $\pi = 0$ represents a *pencil* of $(2p-j)$ -ics, and f_p is a p -ad in some member of the pencil. The given $p-j-1$ roots s_c of g_{p-j-1} , and the known $j+2$ roots t_a determine a $(p+1)$ -secant S_p of N^{2p-1} to which we may regard the above pencil π as attached.

With f_p, f_{p+1}, f_{p+2} related as above we seek new expressions for $(\alpha t)^{2p-1}$ on W_p . According to (2), for every root t_a of λ_{j+2} , $f_{p+1}(t_a)/f_p(t_a) = 0$; and for every root t_b of μ_{2p-j} , $f_{p+1}(t_b)/f_p(t_b) = f_{p+2}(t_b)/f_{p+1}(t_b) = g_{p-j}(t_b)/g_{p-j-1}(t_b)$. Hence the expressions (A), (B) reduce to

$$(D) \quad (\alpha t)^{2p-1} = \sum_{b=j+3}^{b=2p+2} (tt_b)^{2p-1} \cdot g_{p-j}(t_b)/g_{p-j-1}(t_b) \cdot \lambda_{j+2}(t_b) \cdot \mu'_{2p-j}(t_b).$$

There is a linear identity connecting the $(3p-1-j)$ -th powers of the $(3p+1-j)$ linear factors, $2p-j$ of which are factors (tt_b) of μ_{2p-j} ; $j+2$, factors (tt_a) of λ_{j+2} ; and $p-j-1$, factors (ts_c) of g_{p-j-1} . This identity, polarized as to g_{p-j} , yields for the powers of (tt_b) the right member of (D). The remaining powers in the identity then yield the following alternative form of $(\alpha t)^{2p-1}$:

$$(E) \quad -(\alpha t)^{2p-1} = \sum_{a=1}^{j+2} (tt_a)^{2p-1} \cdot g_{p-j}(t_a)/g_{p-j-1}(t_a) \cdot \mu_{2p-j}(t_a) \cdot \lambda'_{j+2}(t_a) \\ + \sum_{c=1}^{p-j-1} (ts_c)^{2p-1} \cdot g_{p-j}(s_c)/\mu_{2p-j}(s_c) \cdot \lambda_{j+2}(s_c) \cdot g'_{p-j-1}(s_c).$$

But, according to (4), for the roots t_a of λ_{j+2} , and the roots s_c of g_{p-j-1} , $g_{p-j}/\mu_{2p-j} = -1/f_p$. Hence

$$(F) \quad (\alpha t)^{2p-1} = \sum_{a=1}^{j+2} (tt_a)^{2p-1}/f_p(t_a) \cdot g_{p-j-1}(t_a) \cdot \lambda'_{j+2}(t_a) \\ + \sum_{c=1}^{p-j-1} (ts_c)^{2p-1}/f_p(s_c) \cdot \lambda_{j+2}(s_c) \cdot g'_{p-j-1}(s_c).$$

The remainder of this article is devoted to a discussion of these formulae (A), \dots , (F).

When the $(p+1)$ -secant S_p of N^{2p-1} , say the $S_p(t_a, s_c)$, is given, the $(\omega t)^{2p+2} = \lambda_{j+2}\mu_{2p-j}$ also being known in advance, the pencil π in (4) is determined, and ∞^1 p -ads f_p of members of the pencil exist, and thus ∞^1 points of W_p are determined. According to (C) such a point is on the

$S_{p-1}(r_e)$, p -secant to N^{2p-1} , and according to (E) it is on the $S_p(t_a, s_c)$. It is thus the unique point common to this S_{p-1} and this S_p .

We have allowed j in (4) to run up to p . The case $j = p$ is quite exceptional, and the case $j = p - 1$ somewhat less so. We examine these two cases in the next section, translating the results obtained to the Kummer K_p . In § 5 and § 6 we return to the other cases.

4. Sections of W_p by F -loci of the p -th and $(p - 1)$ -th kind, and singular spaces $\Sigma^{(p)}$ and $\Sigma^{(p-1)}$ of K_p . If we set $j = p$ in the preceding section so that $p + 2$ branch points of $H_{p^{p+2}}$ are on a line, the pencil π in § 3 (4) reduces to the single member, $\mu_p = -f_p$. Thus the p tangents at O are inflexional, and p of the branch points have run up to O . We find in [1, § 3 (16)] that $H_{p^{p+2}}$ is then represented by any point on the F -locus of the p -th kind, $\pi_{1,2,\dots,p+2}^{(p)} = \pi_{p+3,\dots,2p+2}^{(p)}$, which is the S_{p-1} on the last p points of P_{2p+2}^{2p-1} ; and that this S_{p-1} is mapped by Σ into one of the *singular points*, $\Sigma_{1,2,\dots,p+2;p+3,\dots,2p+2}^{(p)}$ of K_p in S_{2^p-1} . To the various points of this S_{p-1} on W_p there correspond on K_p the various directions about the singular point. Thus the 2^{2p} F -loci of the p -th kind of W_p give rise to the 2^{2p} singular points of K_p , $\Sigma_{i_1 i_2 \dots i_{p+2}; i_{p+3} i_{p+4} \dots i_{2p+2}}^{(p)}$. The F -loci being conjugate under the Cremona group of W_p , the 2^{2p} singular points of K_p are conjugate under the collineation $g_{2^{2p}}$ of K_p , the map by Σ of the Cremona group.

Again, set $j = p - 1$, so that the $p + 1$ branch points t_1, \dots, t_{p+1} of $H_{p^{p+2}}$ are on a line L . The corresponding point of W_p is then on the F -locus, $\pi_{1,\dots,p+1}^{(p-1)}$, the S_p on p_{p+2}, \dots, p_{2p+2} of P_{2p+2}^{2p-1} . Under the de Jonquières involution of order $p + 2$ whose locus of fixed points is $H_{p^{p+2}}$ (which corresponds to $I_{1,2,\dots,2p+2}$ for which W_p is a locus of fixed points) this line L is transformed into a line M on the $p + 1$ branch points, t_{p+2}, \dots, t_{2p+2} , so that the corresponding point of W_p is also on the paired F -locus, $\pi_{p+2,\dots,2p+2}^{(p-1)}$, the S_p on p_1, \dots, p_{p+1} . Thus this point must be on the line common to the two S_p 's. Conversely, any point on this line is on W_p . For, the pencil § 3 (4) is now $\pi = \mu_{p-1} - g_0^2 \lambda_{p+1} = -f_p \cdot g_1$. Since g_1 is a variable linear form as g_0 takes all values, and since (D) is linear in the coefficients of g_1 , the point $(\alpha t)^{2p-1}$ runs over a line, necessarily the line common to the two S_p 's. This line on W_p is mapped by Σ into a rational norm-curve of order p , in the singular space $\Sigma_{1,\dots,p+1;p+2,\dots,2p+2}^{(p-1)}$ which itself has the dimension p [cf. 1 (4), (7)]. In each of the two S_p 's the $p + 1$ points of P_{2p+2}^{2p-1} determine $p + 1$ S_{p-1} 's, F -loci of the p -th kind, each of which meets the line in a point. Such a point maps into a singular point (e. g. $\Sigma_{1,\dots,p;p+1,\dots,2p+2}^{(p)}$) of K_p on N^p . The $2p + 2$ such singular points on N^p are associated with the linear factors g_1 of

the members of the pencil π for the values $g_0 = 0, \infty$. Thus these factors g_1 are the linear factors of $(\omega t)^{2p+2}$ and the $2p+2$ points on N^p have parameters projective to the roots of $(\omega t)^{2p+2}$. Hence

(1) *A pair of F -loci of W_p of the $(p-1)$ -th kind have in common a rational locus which is on W_p itself. These 2^{2p} loci on W_p map into the sections of K_p by its 2^{2p} singular spaces $\Sigma^{(j-1)}$ of dimension p , these sections being rational norm-curves N^p . Each N^p is on $2p+2$ singular points and each singular point is on $2p+2$ N^p 's [cf. 1 (10)]. On each N^p the parameters of the $2p+2$ singular points are projective to the roots of $(\omega t)^{2p+2} = 0$.*

This is the generalization to K_p of the well-known theorem concerning the incidences of singular points and singular conics of the ordinary Kummer surface K_2 in S_3 .

5. Configurations inscribed in the generic curve, the section $[W_p, S_p(t_a, s_c)]$. The section of W_p by the S_p which is $(p+1)$ -secant to N^{2p-1} at the $j+2$ points t_a of P_{2p-1}^{2p-1} and at the $p-j-1$ generic points s_c of N^{2p-1} is not usually irreducible. For sufficiently large values of p , the bisecant lines, or the trisecant planes, etc., will be an F -locus of the p -th kind, and therefore will be on W_p . Then the section of W_p by S_p will contain some of these lines, or planes, etc., as the case may be, which are determined by the $p+1$ crossings of N^{2p-1} and S_p . This part of the section will however contain no generic point of W_p . The significant part of the section is the curve attached to the pencil π of **3** (4). For $j = -2, -1, 0$, and fixed t_a , but variable s_c , these curves cover W_p completely. For $j \geq 1$ they cover completely the section of W_p by an F -locus. We therefore speak of such a curve as the *generic curve of the section* $[W_p, S_p]$.

Reverting to the next to the last paragraph of **3** which states that a point of this curve is cut out on S_p by the S_{p-1} which is p -secant to N^{2p-1} at the p points whose parameters, $f_p = 0$, are a p -ad of a member of the pencil π , we have as an immediate consequence the theorem:

(1) *The pencil π of $(2p-j)$ -ics in **3** (4) is generic except for the peculiarity that one member contains $p-j-1$ double points. This pencil defines on N^{2p-1} a system of ∞^1 $(2p-j)$ -points, each of which determines a COMPLETE figure consisting of $\binom{2p-j}{k+1}$ S_k 's ($k = 0, 1, \dots, 2p-j-1$). The section of these complete figures by $S_p(t_a, s_c)$ yields ∞^1 configurations consisting of $\binom{2p-j}{p+1}$ S_l 's ($l = 0, 1, \dots, p-1$ or $p-j$). The locus of the ∞^1 sets of $\binom{2p-j}{p}$ points of these configurations is the generic curve of the section*

$[W_p, S_p]$. The ∞^1 configurations of S_i 's inscribed in t such that an S_m ($m > l$) is on $\binom{p+m}{m-l}$ S_i 's, and an S_l is

This is the generalization in the direction both of increasing j of a situation which has been observed by Conner in the case $p = 2$, $j = -2$. This case has indicated that the generic plane section of a Weddle is not a generic quartic curve.

6. Involution curves $[W_p, S_p]$; sections of W_p . members of a pencil of binary n -ics are divided into $(n-p)$ -ics, i. e. if $(\alpha t)^n + \lambda(\beta t)^n = f_p \cdot f_{n-p}$, the p -ic substitute an algebraic series (∞^1). Any algebraic curve in correspondence with such a system of p -ics will be called an involution curve. Since the residual p -ic and $(n-p)$ -ic are themselves $I_n^{(p)}$ is also an $I_n^{(n-p)}$. The simplest geometric example of plotting binary p -ics in S_p with reference to a norm is that the coefficients of f_p itself are the coördinates of a point of S_p . We denote this particular type of involution curve by $I_{2p-j}^{(p)}$. The order of this curve is $\binom{n-1}{p-1}$, since an $S_{p-1}(t_1)$ of N points determined by selecting t_2, \dots, t_p from the n -ic contains t_1 .

It is clear from 5 (1) that

(1) *The generic curve of the section, $[W_p, S_p(t_a, s_c)]$, is an involution curve, $I_{2p-j}^{(p)}$.*

We wish to examine this involution curve to see in what cases it is of simple type $[I_{2p-j}^{(p)}, N^p]$, and, in other cases, to find its type. In the formula 3 (E) for a point of this curve we see that if $S_p(t_a, s_c)$ is given, everything in the formula is fixed except a determined constant factor in g_{p-j-1} which runs through the pencil π and may be neglected; and second the coefficients of g_{p-j} , which are the coefficients of a variable $(2p-j)$ -ic of the pencil π . Hence this curve $[W_p, S_p]$ varies in S_p only with the variable coefficients of g_{p-j} . The point is expressed linearly in terms of the $p+1$ rays of R_{2p+1} , which are respectively:

$$\begin{aligned} \pi_a &= (tt_a)^{2p-1}/g_{p-j-1}(t_a) \cdot \mu_{2p-j}(t_a) \cdot \lambda'_{j+1}(t_a) \\ \pi_c &= (ts_c)^{2p-1}/\mu_{2p-j}(s_c) \cdot \lambda_{j+2}(s_c) \cdot g'_{p-j-1}(s_c) \end{aligned}$$

Thus the parametric equation of the point is

$$(3) \quad \sum_{a=1}^{j+2} g_{p-j}(t_a) \cdot \pi_a + \sum_{c=1}^{p-j-1} g_{p-j}(s_c) \cdot \pi_c,$$

and the $p+1$ parameters $g_{p-j}(t_a)$, $g_{p-j}(s_c)$ may be taken as the coördinates of the point in S_p referred to R_{p+1} .

If the coefficients of g_{p-j} are taken as point coördinates in S_{p-j} with reference to an underlying N^{p-j} whose points are given by $(tt_1)^{p-j}$, the dual coördinates may also be taken as the coefficients of a $(p-j)$ -ic in such wise that the incidence condition is the apolarity condition of two $(p-j)$ -ics, the one representing a point and the other an S_{p-j-1} . The hyper-osculating S_{p-j-1} 's of N^{p-j} are then also represented by $(tt_1)^{p-j}$. Thus $g_{p-j}(u_h)$ is the incidence condition of the point g_{p-j} and the S_{p-j-1} of N^{p-j} with parameter u_h . Hence $g_{p-j}(u_1), \dots, g_{p-j}(u_{p-j+1})$ are the point coördinates of the point g_{p-j} referred to the reference figure R_{p-j+1} formed from the S_{p-j-1} 's of N^{p-j} at u_1, \dots, u_{p-j+1} .

The simplest case is $j=0$. In this case the S_{p-j} of g_{p-j} and N^{p-j} may be identified with the $S_p(t_1 t_2 s_1 \dots s_{p-1})$ and the point (3) is merely a transform of the point g_{p-j} on an $[I_{2p-j}^{(p-j)}, N^{p-j}]$. The situation is described by the theorem:

(4) For the case $j=0$, the pencil, $\pi_{2p} = \mu_{2p} - g_{p-1}^2 \lambda_2 = -f_p \cdot g_p$ defines on N^{2p-1} a pencil of $2p$ -points, each $2p$ -point having $2p$ faces (i. e., S_{2p-2} 's on all but one of the $2p$ points). Corresponding to the factorization $\pi_{2p} = f'_1 \cdot g'_{2p-1}$ these faces envelop a rational norm-curve K^{2p-1} , a face of K^{2p-1} and the OPPOSITE point of the $2p$ -point on N^{2p-1} having the same parameter t . The $S_p(t_1 t_2 s_1 \dots s_{p-1})$ is on the $p-1$ faces of the particular $2p$ -point, $g_{p-1}^2 \lambda_2$, which have parameters s_1, \dots, s_{p-1} . Therefore the faces of K^{2p-1} cut S_p in the S_{p-1} 's of a rational norm-curve N^p in S_p with respect to which g_p determines the point $(\alpha t)^{2p-1}$ in $\mathfrak{3}(E)$. The curve $[W_p, S_p(t_1 t_2 s_1 \dots s_{p-1})]$ is the curve $[I_{2p}^{(p)}, N^p]$ of order $\binom{2p-1}{p-1}$ associated with the g_p 's of the above pencil π . On such a section $[W_p, S_p]$ there is an involutorial correspondence set up by the interchange of f_p and g_p .

The only item in this theorem which requires verification is the identification of N^p as the norm-curve with respect to which g_p is plotted. We examine first the $2p$ -ic, μ_{2p} . The face t_4, \dots, t_{2p+2} with parameter t_3 cuts S_p in an S_{p-1} of N^p with parameter t_3 . Thus the faces with parameters t_3, \dots, t_{2p+2} meet in an $S_{p-1}(t_{p+3}, \dots, t_{2p+2})$ which cuts S_p in a point with parameters $g_p = t_3, \dots, t_{p+2}$ with reference to N^p . That this is the point on

W_p determined by g_p in **3** (D) is clear, because, if $g_p(t_3), \dots, g_p(t_{p+2})$ are zero, the point is on the $S_{p-1}(t_{p+3}, \dots, t_{2p+2})$. We examine also the $2p$ -ic, $g_p^2 \cdot \lambda_2 = t_1, t_2, s_1^2, \dots, s_{p-1}^2$. The faces with parameters t_2, s_1, \dots, s_{p-1} meet in an $S_{p-1}(t_1 s_1 \dots s_{p-1})$ which cuts S_p in a point with parameters $g_p = t_2, s_1, \dots, s_{p-1}$ with reference to N^p . But this is the point t_1 on N^{2p-1} itself. Also in **3** (E) for this g_p all the terms vanish except the term in $(tt_1)^{2p-1}$, and thus the point on W_p coincides with the point determined by g_p with respect to N^p . Thus N^p has in common with the norm-curve attached to g_p at least $3p+1$ S_{p-1} 's, and therefore coincides with it.

The case just discussed separates values $j > 0$ from the two values $j < 0$, i. e. $j = -1, -2$. In the latter two cases g_{p-j} is represented on a space of dimension greater than that of S_p ; in the former cases on a space of dimension less than that of S_p . Furthermore in these cases $j > 0$ we are dealing only with points of W_p on an F -locus of the j -th kind. These more nearly resemble the case $j = 0$ and we consider them first.

That the faces of the $2p$ -points in theorem (4) envelop a rational norm-curve K^{2p-1} is well known. So far as we are aware the corresponding theorem, which applies to the cases $j > 0$ and which is given in (5) is new and we incorporate a proof of it.

(5) Let there be given in S_n a norm-curve N^n with parameter t and on it ∞^1 r -points defined by the pencil $(\alpha t)^r + k(\beta t)^r = 0$ [$n+1 \leq r \leq (n+4)/2$]. The S_{r-1} 's determined by two, and therefore by all of these r -points have a common S_{2r-2-n} [$2r-2-n \leq 2$]. Each r -point on N^n has r faces, these being S_{r-2} 's on all but one point of the r -point. The r faces of a particular r -point meet this common S_{2r-2-n} in r S_{2r-3-n} 's and the locus of these S_{2r-3-n} 's in S_{2r-2-n} is a rational norm-curve K^{2r-2-n} which is in face-point correspondence with N^n .

For, it is clear first of all that a particular t_1 determines a particular r -point and that the face of this r -point opposite t_1 is unique. Thus the faces run over a rational locus K . There remains to show that a point in S_{2r-2-n} is on $2r-2-n$ of these faces. If t_1, t belong to the same r -ic of the pencil, they satisfy the symmetric form $(\alpha_1 t_1)^{r-1} (\alpha_2 t)^{r-1} = 0$. For given t_1 , this is the $(r-1)$ -ic which defines the face t_1 , and $(\alpha_1 t_1)^{r-1} (\alpha_2 t)^{r-1} \cdot (tt_1) = 0$ is the r -ic which contains t_1 . If $(\gamma t)^n$ represents, with respect to N^n , a point on S_{2r-2-n} , then $(\gamma t)^n$ is apolar to every r -ic of the pencil, i. e.,

$$(a) \quad (\alpha_1 t_1)^{r-1} (\alpha_2 \gamma)^{r-1} (\gamma t_1) (\gamma t)^{n-r} \equiv 0 \text{ in } t, t_1.$$

This identity (a) can be replaced by the vanishing of the elementary covariants of (a) whose polars figure in the Clebsch-Gordan development, i. e.

$$(b) \quad (\alpha_2 \gamma)^{r-1} (\alpha_1 \gamma)^k (\alpha_1 s)^{r-k-1} (\gamma s)^{n-r-k+1} \equiv 0 \quad (k = 0, \dots, n-r).$$

Similarly the point $(\gamma t)^n$ is on all those faces t_1 for which

$$(c) \quad (\alpha_2 \gamma)^{r-1} (\alpha_1 t_1)^{r-1} (\gamma t)^{n-r+1} \equiv 0 \text{ in } t.$$

In this the elementary covariants of the Clebsch-Gordan development are

$$(\alpha_2 \gamma)^{r-1} (\alpha_1 \gamma)^k (\alpha_1 s)^{r-k-1} (\gamma s)^{n-r-k+1} \quad (k = 0, \dots, n-r+1).$$

But all of these vanish due to (b) except the last whence (c) has the form

$$k \cdot (\alpha_1 t_1)^{2r-2-n} (\alpha_2 \gamma)^{r-1} (\alpha_1 \gamma)^{n-r+1} \cdot (t_1 t)^{n-r+1} \equiv 0 \text{ in } t.$$

Thus $(\gamma t)^n$ is on the $2r-2-n$ faces whose parameters t_1 are given by $(\alpha_1 t_1)^{2r-2-n} (\alpha_2 \gamma)^{r-1} (\alpha_1 \gamma)^{n-r+1} = 0$.

We return now to the pencil **3** (4) for values $j = 1, \dots, p-2$. According to **3** (D), (E) the points of W_p determined by the $(p-j)$ -ics, g_{p-j} , found in members of the pencil π of $(2p-j)$ -ics, lie in the two linear spaces

$$S_{2p-1-j}(t_{j+3}, \dots, t_{2p+2}), \quad S_p(t_1, \dots, t_{j+2}, s_1, \dots, s_{p-j-1}),$$

which meet in a space,

$$(6) \quad S_{p-j}.$$

On N^{2p-1} the pencil π defines ∞^1 $(2p-j)$ -points to which we apply the lemma (5) by means of the transcription:

$$n = 2p-1, \quad r = 2p-j, \quad S_{2r-2-n} = S_{2p-1-2j}.$$

Thus the pencil π on N^{2p-1} determines an $S_{2p-1-2j}$, and the faces of the $(2p-j)$ -points of π cut this $S_{2p-1-2j}$ in the $S_{2p-2-2j}$'s of a rational norm-curve $K^{2p-1-2j}$ in $S_{2p-1-2j}$. The two particular $(2p-j)$ -points, defined by μ_{2p-j} and $g^2_{p-j-1} \cdot \lambda_{j+2}$ have parameters t_{j+3}, \dots, t_{2p+2} and $t_1, \dots, t_{j+2}, s_1^2, \dots, s^2_{p-1-j}$ respectively. Thus the S_{p-j} in (6) is on $S_{2p-1-2j}$. Furthermore, from the particular nature of $g^2_{p-j-1} \cdot \lambda_{j+2}$ in the pencil, this S_{p-j} is on those faces of $K^{2p-1-2j}$ with parameters s_1, \dots, s_{p-j-1} . Thus the faces $K^{2p-1-2j}$ cut S_{p-j} in the S_{p-j-1} 's of a rational norm-curve N^{p-j} in S_{p-j} . Hence

(7) For the cases $j = 1, \dots, p-2$ the pencil $\pi_{2p-j} = \mu_{2p-j} - g^2_{p-j-1} \cdot \lambda_{j+2} = -f_p \cdot g_{p-j}$ defines on N^{2p-1} a pencil of $(2p-j)$ -points whose faces

cut the S_{p-j} (6) in the S_{p-j-1} 's of a rational norm-curve N^{p-j} with respect to which g^{p-j} determines the point of W_p given by 3 (E). The curve $[W_p, S_p(t_1, \dots, t_{j+2}, s_1, \dots, s_{p-j-1})]$ is the curve $[I_{2p-j}^{(p-j)}, N^{p-j}]$ of order $\binom{2p-j-1}{p-j-1}$ associated with the g_{p-j} 's in π .

The identification of N^{p-j} with the norm-curve to which g_{p-j} is attached can be carried out as in the case of the theorem (4).

In the F -space of the j -th kind, $S_{2p-1-j}(t_{j+3}, \dots, t_{2p+2})$ preceding (6), we find $\infty^{p-j-1} \cdot S_{p-j}$'s (6) each containing a curve of the type described in (7) whence

(8) *The non-basic F -loci of the j -th kind meet W_p in manifolds of dimension $p-j$, which are run over by a linear system of ∞^{p-j-1} involution curves. In the case of the linear non-basic F -loci these are involution curves attached to norm-curves.*

The particular cases, $j=p$, $j=p-1$ are discussed in 4.

The cases $j=-1$ and $j=-2$ differ from those just treated in that $(p-j)$ -ics are plotted in a space of dimension respectively one or two greater than that of S_p . In these cases the $p+1$ coefficients $g_{p-j}(t_a), g_{p-j}(s_c)$ in 3 (E) are the $p+1$ coördinates of a point in S_{p-j} when this point is projected upon an S_p from the point π_0 ($j=-1$), or line π_1 ($j=-2$), in which the hyperosculating spaces t_a, s_c of N^{p-j} meet. Moreover the reference R_{p+1} in S_p to which these coördinates refer after the projection has vertices at t_a, s_c where S_p cuts N^{2p-1} . Hence

(9) *The generic curves, $[W_p, S_p(t_1, \dots, t_{j+2}, s_1, \dots, s_{p-j-1})]$, defined by the pencil π for $j=-1, -2$ are projections of the involution curve $[I_{2p-j}^{(p-j)}, N^{p-j}]$ of order $\binom{2p-j-1}{p-j-1}$ defined by g_{p-j} -ics of π with reference to N^{p-j} from the linear space π_{j-1} in which the hyperosculating spaces of N^{p-j} with parameters t_a, s_c meet.*

We may therefore make the general statement that

(10) *The curves cut out on W_p by $(p+1)$ -secant S_p 's of N^{2p-1} are either involution curves attached to norm-curves [cf. (1), (7)], or they are projections of such curves [cf. (9)].*

An entirely different aspect of these curves is brought out by the formula 3 (F) which we have not as yet used. This formula contains the coefficients of the factor f_p residual to g_{p-j} in the pencil π . Let N^p be a norm-curve in S'_p with reference to which the p -ics f_p are plotted. Then $f_p(t_a), f_p(s_c)$ are

coördinates in S'_p with respect to the reference figure R'_{p+1} whose $p+1$ spaces hyperosculate N'^p at the points t_a, s_c of N'^p . As before $1/f_p(t_a), 1/f_p(s_c)$ are coördinates in S_p with respect to the reference figure R_{p+1} whose $p+1$ vertices are the points t_a, s_c of N^{2p-1} . Since f and $1/f$ are related by a Cremona transformation, we have the theorem:

(11) *Let the p -ics, f_p , of the pencil, $\pi \equiv \mu_{2p-j} - g_{p-j-1}^2 \cdot \lambda_{j+2} \equiv -f_p \cdot g_{p-j}$, plotted with respect to the norm-curve N'^p in S'_p , define the involution curve $(I_{2p-j}^{(p)}, N'^p)$ of order $\binom{2p-j-1}{p-1}$. Let R'_{p+1} be the reference figure in S'_p whose S'_{p-1} 's hyperosculate N'^p with parameters t_a, s_c . Let S_p be the $(p+1)$ -secant space of N^{2p-1} at the points R_{p+1} with parameters t_a, s_c . Then the regular Cremona transformation of order p with direct and inverse F -points at the points of R'_{p+1}, R_{p+1} respectively transforms the above involution curve into the generic curve of the section of W_p by S_p .*

A version of the inverse Cremona transformation is obtained by taking the canonizant of $(at)^{2p-1}$ in $\mathbf{3}(E)$. This canonizant, according to $[\mathbf{3}(1), (C)]$, is f_p itself. But the canonizant of $\mathbf{3}(E)$ is of degree p in the coefficients of g_{p-j} .

7. Applications to W_2, W_3, W_4 . The generic section of W_2 , the Weddle surface in S_3 , by an S_2 on the points s_1, s_2, s_3 of the cubic curve N^3 on the six nodes P_6^3 of W_2 is a quartic curve which, according to the Morley-Conner theorem, generalized in $\mathbf{5}(1)$, contains ∞^1 configurations $(15_3, 20_4)$. These are the sections of the complete 6-points determined on N^3 by the pencil, $\pi = \mu_6 - f_3^2 = -f_2 f_4, f_3$ having roots s_1, s_2, s_3 .

The general theorems of the preceding section, $\mathbf{6}(9)$ and $\mathbf{6}(11)$, present this curve under two new aspects. The first aspect is that of tetrads, f_4 , of members of the pencil π . If these are plotted as points in S_4 with reference to an N^4 , the generic sextic of the pencil determines 15 points of the involution curve $(I_6^{(4)}, N^4)$, of which 10 are in a particular osculating S_3 of N^4 . For the particular sextic, f_3^2 , these 15 points comprise three of type, $b_1 = s_2^2 s_3^2$; and three of type, $a_1 = s_1^2 s_2 s_3$, each counting four times. The latter three points, a_1, a_2, a_3 , are on the line of intersection of the three osculating spaces s_1, s_2, s_3 of N^4 ; and they are double points of $(I_6^{(4)}, N^4)$. For, the osculating space s_1 contains a_1 counting four times, and a_2, a_3 each counting twice. But also the osculating plane s_1^2 contains a_1 counting four times, and a_1 is thus a node with tangents in the plane s_1^2 . Hence $(I_6^{(4)}, N^4)$ of order 10, projected from the line on its three nodes a_i , yields the quartic plane section of W_2 .

The second aspect is that of pairs f_2 of members of the pencil π . We take then in a plane π' a norm-conic N'^2 and the six-lines π circumscribed

about it, each six-line contributing 15 points on the involution curve $(I_6^{(2)}, N'^2)$ of order 5. The particular member, f_3^2 , contributes a circumscribed triangle s_1, s_2, s_3 of N'^2 , the 15 points being the three points $c_1 = s_1^2, c_2 = s_2^2, c_3 = s_3^2$ of contact; and the three vertices, $d_1 = s_2s_3, d_2 = s_1s_3, d_3 = s_1s_2$, each counting four times. The five points on the line s_3 of N'^2 are s_3^2 on N'^2 and d_1, d_2 each counting twice. Hence d_1, d_2, d_3 are nodes of the involution curve. According to 6.(11) the section of W_2 is the transform of this quintic $(I_6^{(2)}, N'^2)$ by the quadratic transformation A_{123} with F -points at its nodes d_1, d_2, d_3 . The section is therefore a quartic curve. If t_1, t_2, t_3 is a triad of any member of π , the vertices of the two circumscribed triangles, t_1, t_2, t_3 and s_1, s_2, s_3 , of N'^2 are on a conic. This conic is transformed by A_{123} into a line and thus we find on the section of W_2 the inscribed Morley-Conner configurations.

The conic N'^2 itself passes by A_{123} into the tri-cuspidal quartic curve whose envelope is the rational cubic of lines cut out on the plane of the section by the osculating planes of N^3 . The cusp triangle cut out on the plane by N^3 is the triangle of inverse F -points. These interesting connections, and the particular cases which arise from sections of W_2 by planes on one or two nodes of W_2 , are deserving of further study.

In passing to the consideration of W_3 in S_5 and W_4 in S_7 we utilize only the second aspect mentioned above, since we are interested primarily in the order of these loci. We consider the section of W_3 by the S_3 quadrisecant to N^5 at $f_4 = s_1, s_2, s_3, s_4$ and the associated pencil $\pi = \mu_8 - f_4^2 = -f_3 \cdot f_5$. The triads f_3 of members of the pencil are mapped by points f_3 in the space S'_3 of a cubic curve N'^3 , which lie on the involution curve $(I_8^{(3)}, N'^3) = K^{21}$ of order 21. If t_1, \dots, t_8 is a generic octavic of the pencil π , the osculating plane t_1 of N'^3 cuts K^{21} in the 21 points $t_1t_2t_3, \dots, t_1t_7t_8$. The axis t_1t_2 of N'^3 cuts K^{21} in 6 points $t_1t_2t_3, \dots, t_1t_2t_8$. The osculating plane t_1 of N'^3 is cut by the planes of N'^3 in the lines of a conic $K^2(t_1)$ which touches the tangent to N'^3 at the point t_1 of N'^3 . The 7 axes cut out on the plane t_1 of N'^3 by planes t_2, \dots, t_7 envelop $K^2(t_1)$ and their 21 meets are on K^{21} .

The pencil π has no other peculiarity than that it contains one square member, f_4^2 . For this member the inscribed 8-plane of K^{21} collapses to a 4-plane. The seven axes enveloping $K^2(s_1)$ are now the tangent to N'^3 at s_1 ; and the three axes s_1s_2, s_1s_3, s_1s_4 , each counting twice. The tangent, or axis s_1^2 , meets these three axes in points $s_1^2s_2, s_1^2s_3, s_1^2s_4$. Thus these three points are contacts of a tritangent line of K^{21} , namely, the tangent to N'^3 at s_1 . The 6 points of K^{21} on the axis s_1s_2 are now $s_1s_2s_3, s_1s_2s_4$, each counting twice; and $s_1^2s_2$ and $s_1s_2^2$ (the contact of the tangent s_2 of $K^2(s_1)$), each counting once. The plane s_1 cuts K^{21} in points $d_2 = s_1s_3s_4, d_3 = s_1s_2s_4, d_4 = s_1s_2s_3$, each

counting four times; the points $s_1^2s_2$, $s_1^2s_3$, $s_1^2s_4$, each counting twice; and the points $s_1s_2^2$, $s_1s_3^2$, $s_1s_4^2$, each counting once. Since the point $d_1 = s_1s_2s_3$ is four-fold on each of the three planes s_1 , s_2 , s_3 containing it, it is a four-fold point of K^{21} . Thus the vertices d_i of the tetrahedron formed by the planes s_i of N^3 are four-fold on K^{21} , and the edges s_1s_j of this tetrahedron meet K^{21} in the two further points $s_i^2s_j$, $s_1s_j^2$.

According to 6 (11) the section of W_3 by the $S_3(s_1 \cdots s_4)$ is the transform of K^{21} by the regular cubic transformation A_{1234} with F -points at d_1, \dots, d_4 on N^3 and with inverse F -points at s_1, \dots, s_4 on N^5 . Since the order of the transform of a curve by A_{1234} is reduced by two for each branch through an F -point, and by one for each crossing of an F -line joining two F -points, the transform of K^{21} by A_{1234} has the order $3 \cdot 21 - 2 \cdot 4 \cdot 4 - 1 \cdot 6 \cdot 2 = 19$. This transform L^{19} has triple points at s_1, \dots, s_4 . Due to the contacts of K^{21} with the tangent s_1^2 in S'_3 the tangents of L^{19} at the triple point s_1 are the three lines s_1s_2 , s_1s_3 , s_1s_4 . Hence, making use also of 5 (1),

(1) *The section of W_3 in S_5 by a quadri-secant $S_3(s_1 \cdots s_4)$ of N^5 is a curve of order 19 with triple points at the four points s_i on N^5 and tangents s_1s_j at the triple point s_i . This curve contains ∞^1 inscribed configurations, each consisting of 56 points, 70 lines, and 56 planes with the following incidences: each point is on 5 lines and 10 planes, each plane is on 5 lines and 10 points, and each line is on 4 points, 16 lines, and 4 planes.*

The inverse transformation, A_{1234}^{-1} , transforms L^{19} back into K^{21} , the reduction being $3 \cdot 19 - 4 \cdot 3 \cdot 2 - 12 \cdot 1 = 21$, where the reduction $12 \cdot 1$ arises from the contacts of the edges of the tetrahedron s_i at the triple points. We have thus the confirmation of an earlier result [cf. ¹, (81)], and the more precise information given by

(2) *The W_3 in S_5 has the order 19 and has the triple curve N^5 on P_8^5 and therefore also triple lines p_ip_j . The tangent cone at a point s of N^5 contains the bisecants of N^5 on s .*

This explains the behavior of a trisecant plane of N^5 which must cut W_3 in 19 points. Through any point of W_3 there is just one such trisecant plane, namely, the plane r_1, r_2, r_3 of 2 (17b). This plane cuts W_3 in the four ordinary points obtained by the variation of the signs of z_{r_i} . The intersections with the triple curve N^5 at the points r_i account for 9, and the contacts of the plane with W_3 along the line r_ir_j at r_i and r_j account for the remaining 6, of the 19 intersections. It is these four ordinary points in the trisecant plane $s_2s_3s_4$ which pass by A_{1234}^{-1} into the four-fold point of K^{21} at the point d_1 .

We examine next the intersection of W_3 by the quadrisecant S_3 of N^5 on s_1, s_2, s_3 roots of $g_3 = 0$, and t_1 , a point of P_8^5 . Since the cone of lines on t_1 to points of N^5 is an F -locus of the third kind lying on W_3 , the three lines from t_1 to s_1, s_2, s_3 will separate from the intersection leaving a curve L^{16} , which is associated with the pencil $\pi = \mu_7 - g_3^2 \lambda_1 = -f_3 \cdot g_4$. We first examine the involution curve, $(I_7^{(3)}, N'^3) = K^{15}$, determined by the triads f_3 of the pencil with reference to an N'^3 in S'_3 . The four planes s_1, s_2, s_3, t_1 of N'^3 form a tetrahedron with respective opposite vertices d_1, d_2, d_3, d_4 . By considering the multiplicities of the 15 points of K^{15} on each of the four planes, and of the 5 points of K^{15} on each of the six edges, of this tetrahedron, we find that: (a) the point d_4 is four-fold, and the points d_i are double, on K^{15} ; (b) the point $s_i^2 s_j$ is on K^{15} , and the tangent to K^{15} at the point is in the plane s_i ; (c) the point $s_i^2 t_1$ on the edge $s_i t_1$ is on K^{15} with tangent neither in the plane s_i nor in the plane t_1 ; and (d) the pairs of tangents of K^{15} at the points d_i are in the plane t_1 [$i, j = 1, 2, 3$]. The transformation A_{1234} with F -points at d_1, d_2, d_3, d_4 and respective inverse F -points at s_1, s_2, s_3, t_1 in the S_3 -section of W_3 transforms K^{15} into L^{16} with multiplicities 2 at s_i and 6 at t_1 , due to the 4 F -points at d_1, \dots, d_4 and the 9 F -points on the edges. Due to (b), the tangents of L^{16} at the node s_i are $s_i s_j$ and $s_i s_k$. Due to (c), the line $s_i t_1$ on W_3 cuts L^{16} at a point distinct from s_i and t_1 . Due to (d), the tangents of L^{16} at t_1 are found two in each of the three planes $t_1 s_i s_j$. We have thus confirmed the 9-fold character of the F -point p_1 of W_3 [cf. ¹, (81)]. The trisecant plane $t_1 s_1 s_2$ cuts L^{16} in 8 points at t_1 , in 3 points at s_1 and at s_2 , and in one point on each of $t_1 s_1$ and $t_1 s_2$. These points all are on F -loci of the second or third kinds. Thus the F -locus of the first kind, $\pi_2^{(1)} \dots_8$, the locus of S_2 's on t_1 and two variable points of N^5 meets W_3 in no significant locus. It is indeed paired with $\pi_1^{(1)}$, the directions about t_1 , and thus the two F -loci meet each other and W_3 only in the directions about t_1 on W_3 , apart from intersections on F -loci of higher kinds.

The intersection of W_3 by the quadrisecant S_3 of N^5 on s_1, s_2 , roots of $g_2 = 0$, and on t_1, t_2 , points of P_8^5 , is associated with the pencil, $\pi = \mu_6 - g_2^2 \lambda_2 = -f_3 \cdot g_3$. This intersection must contain as a part the four bisecants $t_i s_j$, and it must contain the F -line $t_1 t_2$ as a five-fold intersection, three-fold due to the multiplicity of $t_1 t_2$ and twice as a contact, this being due to the order of the residual intersection L^{10} associated with the triads g_3 of the pencil [cf. § (4)]. The triads f_3 of this pencil determine in S'_3 with reference to N'^3 an involution curve, $(I_6^{(3)}, N'^3) = K^{10}$. The four planes of N'^3 with parameters t_1, t_2, s_1, s_2 have respective opposite vertices d_1, d_2, e_1, e_2 . The special sextic $g_2^2 \lambda_2$ requires that K^{10} have the following properties: (a) the points

d_1, d_2 are nodes, e_1, e_2 simple points, of K^{10} , the edge e_1e_2 being tangent at e_1, e_2 ; (b) the edge s_1s_2 meets K^{10} in two simple points $s_1^2s_2, s_1s_2^2$, the tangent at $s_1^2s_2$ lying in the plane s_1 ; and (c) the edge s_it_j meets K^{10} in the point $s_i^2t_j$. The transformation A_{1234} with F -points at d_1, d_2, e_1, e_2 and respective inverse F -points at t_1, t_2, s_1, s_2 on N^5 transforms K^{10} into L^{10} with multiplicities 2, 2, 1, 1 at t_1, t_2, s_1, s_2 , due to the multiplicities of K^{10} at the four F -points and at the eight F -points of the second kind on the edges. Due to (b), the edge s_1s_2 touches L^{10} at s_1 and s_2 ; and, due to (c), L^{10} is crossed by s_it_j . Due also to the contacts at e_1, e_2 , the tangents to L^{10} at the points on the edge t_1t_2 are respectively on the planes containing this edge. Thus L^{10} and K^{10} are each related to their respective tetrahedra in precisely the same way as might be inferred directly from the mutual relation of f_3, g_3 to π . The most striking new property that has appeared is:

(3) *A quadrisecant S_3 of N^5 with two crossings at t_1, t_2 has a contact of second order with W_3 along the line t_1t_2 in addition to the expected triple intersection.*

We examine finally the intersection of W_3 by the quadrisecant S_3 of N^5 on s_1 and on t_1, t_2, t_3 , points of P_8^5 . This is made up of the plane $t_1t_2t_3$ and a residual curve part of which is made up of t_1s_1 and of the lines t_it_j to a certain multiplicity. The significant part of the curve is associated with the pencil $\pi = \mu_5 - g_1^2\lambda_3 = -f_3 \cdot g_2$. In this case we observe at once that the duads g_2 of the pencil determine a Lüroth quartic curve in the plane S_2 cut out on S_3 by the $S_4(p_4, \dots, p_8) = \pi_{123}^{(1)}$ [cf. 6 (6)]. Thus

(4) *The 2-way of order 9 [cf. ¹, p. 489] cut out on W_3 by the F -locus $\pi_{123}^{(1)}$, the ten planes $p_4p_5p_6, \dots, p_6p_7p_8$ being disregarded, contains a linear system of planar Lüroth quartics.*

We conclude with a discussion of the intersection of W_4 in S_7 with an S_4 5-secant to N^7 at points s_1, \dots, s_5 of N^7 given by $f_5 = 0$. In the case of W_4 , the bisecant locus of N^7 is an F -locus of the fourth kind contained simply on W_4 , whence the intersection in question will contain as a part the 10 bisecant lines of N^7 joining the points s . The remaining significant intersection is associated with the pencil, $\pi = \mu_{10} - f_5^2 = -f_4 \cdot f_6$. We plot the tetrads f_4 of members of this pencil in S_4' with reference to an N'^4 to obtain the involution curve $(I_{10}^{(4)}, N'^4) = K^{84}$. The degenerate member f_5^2 contributes five S_3 's of N'^4 with parameters s_1, \dots, s_5 whose five vertices are $d_i = s_js_ks_ls_m$. An examination of the 84 intersections of these five S_3 's with K^{84} yields the following results: (a) the five points d_i are 8-fold on K^{84} ;

(b) the edge $d_i d_j$ touches K^{84} at three points $d_{ij,k} = s_k^2 s_i s_m$; and (c) the plane $d_i d_j d_k$ cuts K^{84} at $d_{ijk} = s_i^2 s_m^2$. The quartic transformation A_{12345} with F -points at d_1, \dots, d_5 , and inverse F -points at the points s_1, \dots, s_5 on N^7 in S_7 transforms K^{84} into the section L of W_4 by the 5-secant S_4 . Due to the 5 8-fold F -points d_i , the 30 repeated F -points on edges $d_i d_j$, and the 10 F -points on planes $d_i d_j d_k$, the order of the transform is $4 \cdot 84 - 5 \cdot 8 \cdot 3 - 30 \cdot 2 \cdot 2 - 10 \cdot 1 = 86$, and the section L^{86} has 12-fold points at the five points s_i . On adding the 10 lines of this five-point we have the theorem:

(5) W_4 in S_7 has the order 96, and it contains N^7 and the 45 lines $p_i p_j$ as 16-fold curves.

The point $d_{ij,k}$ on K^{84} by virtue of (b) passes into a direction on L^{86} on the trisecant plane $s_k s_i s_m$ of N^7 at s_k , this direction counting doubly. The six trisecant planes on s_i thus account for the multiplicity 12 of L^{86} at s_i . The point d_{ijk} on K^{84} , by virtue of (c), passes into a point of L^{86} on the line $s_1 s_2$. Thus L^{86} is transformed back into K^{84} since $4 \cdot 86 - 5 \cdot 12 \cdot 3 - 10 \cdot 2 - 30 \cdot 2 \cdot 1 = 84$. The 8-fold point of K^{84} at d_1 arises from the eight generic points of W_4 in the quadric-secant S_3 on s_2, \dots, s_5 [cf. 2 (17c)]. The remaining 78 intersections of this S_3 with L^{86} are accounted for by the 48 at s_2, \dots, s_5 , by the 6 on the edges $s_2 s_3, \dots$, and by the $4 \cdot 3 \cdot 2$ directions at s_2 in the planes $s_2 s_3 s_4, \dots$.

REFERENCES.

¹ A. B. Coble, "A generalization of the Weddle Surface, . . .," *American Journal of Mathematics*, vol. 52 (1930), pp. 439-500.

² A. B. Coble, "Hyperelliptic functions and irrational binary invariants I," *American Journal of Mathematics*, vol. 54 (1932), pp. 425-452.

³ A. B. Coble, "Algebraic geometry and theta functions," *Colloquium Lectures of the American Mathematical Society*, vol. 10, New York (1929).

⁴ F. Morley and J. R. Conner, "Plane sections of the Weddle surface," *American Journal of Mathematics*, vol. 31 (1909), pp. 263-270.

⁵ A. B. Coble, "The geometry of the Weddle manifold, W_4 ," *Bulletin of the American Mathematical Society*, vol. 41 (1935), April number.

A THEORY OF POSITIVE INTEGERS IN FORMAL LOGIC.*

PART II.

By S. C. KLEENE.

15. **Formal definition: initial values, induction.** If L is an intuitive function which associates well-formed expressions $L(x_1, \dots, x_n)$ with n -tuples (x_1, \dots, x_n) of well-formed expressions, then L shall be said to be *defined (formally)* by L if $L(x_1, \dots, x_n) \text{ conv } L(x_1, \dots, x_n)$ for each set (x_1, \dots, x_n) for which L is defined. By the "definition" of a function which correlates intuitive mathematical objects, we shall mean the definition of the function which correlates the corresponding well-formed formulas, in case corresponding formulas have been designated. By the "definition" of a sequence A_1, A_2, A_3, \dots , we shall mean the definition of a function L whose values for the arguments $1, 2, 3, \dots$ are A_1, A_2, A_3, \dots , respectively. That is, A_1, A_2, A_3, \dots shall be defined (formally) by L , if $L(i) \text{ conv } A_i$ ($i = 1, 2, 3, \dots$).

Closely connected with the formal theory of this paper, there is an intuitive theory concerning the formal definition of the functions involved. For the preceding sections, this may be summarized by the following theorem, each part of which can be established, either directly, with the aid of the first, or by means of considerations used above in formal proofs.†

15I. Suppose that x and y are given positive integers of intuitive logic.
a. $x \text{ conv } \lambda f a \cdot f(\dots x \text{ times } \dots f(a) \dots)$. b. If $x + y = z$, $x + y \text{ conv } z$.
c. If $xy = z$, $xy \text{ conv } z$. d. $F^x(A) \text{ conv } F(\dots x \text{ times } \dots F(A) \dots)$.
e. $I^x \text{ conv } I$; $I(A) \text{ conv } A$. f. If $x^y = z$, $x^y \text{ conv } z$. g. If $x > y$ and $x - y = z$, $x - y \text{ conv } z$; if $x \leq y$, $x - y \text{ conv } 1$. h. If $x \leq y$, $\min(x, y) \text{ conv } \min(y, x) \text{ conv } x$. i. If $x \geq y$, $\max(x, y) \text{ conv } \max(y, x) \text{ conv } x$.
j. $1 \circ 1 \text{ conv } 1 \circ 2 \text{ conv } 2 \circ 1 \text{ conv } 1$; $2 \circ 2 \text{ conv } 2$. k. If $x \leq y$, $\epsilon_y^x \text{ conv } 1$; if $x > y$, $\epsilon_y^x \text{ conv } 2$. l. If $x \neq y$, $\delta_y^x \text{ conv } 1$; if $x = y$, $\delta_y^x \text{ conv } 2$.‡

* Part I appeared in this Journal, vol. 57 (1935), pp. 153-173.

† 15I is stated with the aid of the convention that if n represents a positive integer of intuitive logic, then n shall represent the corresponding positive integer $S(\dots n - 1 \text{ times } \dots S(1) \dots)$ of our formal theory.

‡ This theorem includes the assertion that the intuitive functions $x + y$, xy , x^y , $x - y$, $\min(x, y)$, $\max(x, y)$ are definable (for positive integral arguments and values). Also, constant and identity functions of positive integers are definable: If n, x_1, \dots, x_n are given positive integers, then $G(n, A, x_1, \dots, x_n) \text{ conv } A$ and $\mathfrak{I}_{ni}(\mathfrak{I}_n(x_1, \dots, x_n)) \text{ conv } x_i$, where $\mathfrak{I}_n \rightarrow \lambda p_1 \dots p_n f_1 \dots f_n a \cdot f_1 p_1(\dots f_n p_n(a))$ and $\mathfrak{I}_{ni} \rightarrow \lambda p f \cdot p(I, \dots, i - 1 \text{ times } \dots, I, f, I, \dots, n - i \text{ times } \dots, I)$ (cf. §§ 7, 8).

The remainder of this paper is devoted to further developments of the theories of formal definition and of formal proof in conjunction with each other.

15II. *A necessary condition that a function of positive integers, the values of which are well-formed expressions, be definable is that all the values have the same free symbols.*

This is a consequence of C5VI.

15III(k). *If A_1, \dots, A_k, F have the same free symbols, then the sequence $A_1, \dots, A_k, F(1), F(2), \dots$ is definable by a formula L such that $N(X) \vdash' L(k + X) = F(X)$.**

Proof. If A and B have the same free symbols, then, by C7I, there exists a formula B such that $B(1) \text{ conv } \lambda n \cdot I^n(A)$ and $B(2) \text{ conv } F$. Let $L \rightarrow \lambda n \cdot B(\min(2, n), n - 1)$.† Then it is clear from 15Ie, g, h that L defines $A, F(1), F(2), \dots$. Also $N(X) \vdash' L(1 + X) = F(X)$, since, assuming $N(X)$, we have $L(1 + X) \text{ conv } B(\min(2, S(X)), S(X) - 1)$, $= B(2, X)$ (13.2, 12.4, 11.2), $\text{conv } F(X)$. Thus 15III(1) is established. Moreover 15III($k + 1$) is a consequence of 15III(k) and 15III(1).‡ Thus 15III(k) is established by an intuitive induction with respect to k .

COROLLARY. *If $A_{i_1} \dots i_n$ ($i_1, \dots, i_n = 1, \dots, k$) have the same free symbols, then a formula L can be found such that $L(i_1, \dots, i_n) \text{ conv } A_{i_1} \dots i_n$.*

This follows from 15III by induction with respect to n , since, given the hypothesis with $n + 1$ replacing n , we can, by using the corollary as stated, find k formulas L_{i_1} such that $L_{i_1}(i_2, \dots, i_{n+1}) \text{ conv } A_{i_1} \dots i_{n+1}$, and then by 15III find an L such that $L(i_1) \text{ conv } L_{i_1}$.

15IV. *If the free symbols of F are included among those of A , then the sequence $A, F(A), F(F(A)), \dots$ is definable by a formula L such that $N(X) \vdash' L(S(X)) = F(L(X))$.*

* For the notation $\vdash' =$ see the last paragraph of § 2.

† When a heavy-typed letter represents occurrences of a proper symbol in a formula, we shall suppose the symbol to be one whose only occurrences in the formula are those represented by the occurrences of the letter, unless the contrary is implied by the conventions (1) and (2) of § C3. Thus n is here supposed to be distinct from the proper symbols of A and B , but in " $\lambda n \cdot M$ " n must occur in M as a free symbol in order that M and $\lambda n \cdot M$ be well-formed.

‡ Using the fact that if L' defines $A_2, \dots, A_{k+1}, F(1), F(2), \dots$, then L' has the same free symbols as A_2, \dots, A_{k+1} and F (cf. C5VI). Similarly below.

the assumption), $\text{conv } \lambda \mu f a \cdot \mu(i+3, \mathfrak{R}(i+1, f, f, a), \mathfrak{R}(i, f, f, a), \dots, \mathfrak{R}(1, f, f, a), f(1, a), a)$. Hence, by induction with respect to i , $\mathfrak{R}(i+1) \text{ conv } \lambda \mu f a \cdot \mu(i+2, \mathfrak{R}(i, f, f, a), \dots, \mathfrak{R}(1, f, f, a), f(1, a), a)$. By 15III, there can be found an expression L which defines $A_1, A_2, \mathfrak{R}(1, F, F, A_1), \mathfrak{R}(2, F, F, A_1), \dots$. Then $L(3) \text{ conv } \mathfrak{B}(F, F, A_1), \text{ conv } A_3$. Assuming that $L(j) \text{ conv } A_j$ ($j = 1, \dots, i+2$), then $L(i+3) \text{ conv } \mathfrak{R}(i+1, F, F, A_1), \text{ conv } \{\lambda \mu f a \cdot \mu(i+2, \mathfrak{R}(i, f, f, a), \dots, \mathfrak{R}(1, f, f, a), f(1, a), a)\}(F, F, A_1)$ (as shown above), $\text{conv } F(i+2, L(i+2), \dots, L(3), L(2), L(1))$, $\text{conv } F(i+2, A_{i+2}, \dots, A_1)$ (by hyp.), which is A_{i+3} . Hence, by induction, $L(i) \text{ conv } A_i$ ($i = 1, 2, \dots$).

In 15III-15VII, the expressions F, A, A_1, \dots, A_k of the hypotheses may be replaced by any definable functions of given numbers of positive integers. For example, 15V can be generalized thus: *If the free symbols of F are included among those of A , then a formula L can be found such that $L(x_1, \dots, x_n, y_1, \dots, y_m, 1) \text{ conv } A(x_1, \dots, x_n)$ and $L(x_1, \dots, x_n, y_1, \dots, y_m, i+1) \text{ conv } F(y_1, \dots, y_m, i, L(x_1, \dots, x_n, y_1, \dots, y_m, i))$ ($x_1, \dots, y_m = 1, 2, \dots$). For if $A' \rightarrow I^{b_1}(\dots I^{b_m}(A(a_1, \dots, a_n)) \dots)$ and $F' \rightarrow I^{a_1}(\dots I^{a_n}(F(b_1, \dots, b_m)) \dots)$, where a_1, \dots, b_m are distinct proper symbols, then, by 15V, there exists an expression L' which defines $A', F'(1, A'), \dots$, and we may take for L the function $\lambda r_1 \dots s_m p \cdot \{\lambda a_1 \dots b_m \cdot L'(p)\}(r_1, \dots, s_m)$. Any of the parameters x_1, \dots, y_m of the sequence defined by $L(x_1, \dots, y_m)$ may be equated, since a function obtained from a definable function by equating (or interchanging) a pair of variables is definable (provided the domains of the two variables are the same). For if L defines $L(x, y)$, then $\lambda x \cdot L(x, x)$ defines $L'(x)$ where $L'(x) = L(x, x)$ and $\lambda xy \cdot L(y, x)$ defines $L''(x, y)$ where $L''(x, y) = L(y, x)$; and similarly for functions of more variables. A function obtained from a definable function by substituting for a certain variable a definable function of other variables is definable (provided the domain of the replaced variable contains the domain of values of the substituted function).*

It is clear from the foregoing that every function recursive in the limited sense of Gödel (1931)* is definable, if we use $\lambda fx \cdot f(x), S(\lambda fx \cdot f(x)), S(S(\lambda fx \cdot f(x))), \dots$ as formulas for the numbers $0, 1, 2, \dots$, resp. (thus going over from our theory of positive integers to a like theory of natural numbers), or if we replace natural numbers by positive integers in Gödel's theory. In either case Gödel's Theorems I-IV provide a convenient means

* Kurt Gödel, "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I," *Monatshefte für Mathematik und Physik*, vol. 38 (1931), pp. 173-198. Cf. p. 179.

for showing that various functions, such as quotient, remainder, highest common factor, n -th prime number, are definable.*

It is also true that functions recursive in various more general senses may be defined formally.†

In some situations in which one of the above methods can be used a special device may be more expeditious.

Situations which do not come precisely within the scope of any one of the theorems of this and the following sections may often be dealt with by using several of them and by employing supplementary devices. As a general method of procedure, when it is not at once evident how to define a sequence K_1, K_2, \dots , we attempt to find another sequence K'_1, K'_2, \dots and a J such that $J(K'_1) \text{ conv } K_1, J(K'_2) \text{ conv } K_2, \dots$ and to define K'_1, K'_2, \dots ; or, more generally, to find and define two other sequences K'_1, K'_2, \dots and K''_1, K''_2, \dots such that $K''_1(K'_1) \text{ conv } K_1, K''_2(K'_2) \text{ conv } K_2, \dots$.

In case there is given a recursive situation like that in one of our theorems but with the function relating the members of the sequence in intuitive logic, the difficulty of finding a function F of the formal logic relating the members may often be evaded by the introduction into the terms of the sequence of an extra bound symbol on which a substitution can be made which transforms any member of the sequence K'_1, K'_2, \dots thus obtained into the next member.

Given a positive integer n , let n_0 denote n^n , and n_{k+1} denote $(\dots(n_k)_k \dots)_k$ (n_k subscripts). n_n as a function of n is defined formally by \mathfrak{B} if $\mathfrak{B} \rightarrow \lambda n \cdot [\lambda \rho m \cdot \rho^{\rho(m)}(m)]^n (\lambda r \cdot r^r, n)$. It is amazing that such a brief formula as $\mathfrak{B}(\mathfrak{B})$ should have so long a normal form (cf. § C5).

16. Finite sums and products. Let $\mathfrak{f} \rightarrow \lambda \pi \rho f m \cdot \rho(f, m) + f(m + \pi)$. By 15V, the sequence $1, \mathfrak{f}(1, 1), \mathfrak{f}(2, \mathfrak{f}(1, 1)), \dots$ is definable by a formula \mathfrak{E}

* In the first case, it should be noted at the outset that sum, product, difference, etc., are definable in the resulting theory of natural numbers.

In the second case, the absence of 0 causes no difficulty in proving Gödel's I-IV (as modified in statement by the change from natural numbers to positive integers), since 0 may be used to multiply 1's and 2's as 0's and 1's, respectively (cf. 15Ij).

† As an example, given formulas F and G having the same free symbols, to obtain a formula H such that $H(1, n) \text{ conv } F(n)$, $H(m+1, 1) \text{ conv } G(m)$, and $H(m+1, n+1) \text{ conv } H(m, H(m+1, n))$ ($m, n = 1, 2, \dots$), we may use 15III-15V, according to which formulas L , \mathfrak{R} , and \mathfrak{S} can be found such that $L(1) \text{ conv } F$, $L(2) \text{ conv } G$, $\mathfrak{R}(1) \text{ conv } \lambda h x y l \cdot h(1, Iy, I, l(2, x))$, $\mathfrak{R}(n+1) \text{ conv } \lambda h x y l \cdot h(\mathfrak{R}(n, h, x, y-1, l), l)$, $\mathfrak{S}(1) \text{ conv } \lambda y l \cdot l(1, y)$, and $\mathfrak{S}(m+1) \text{ conv } \lambda y \cdot \mathfrak{R}(y, \mathfrak{S}(m), m, y)$, and let $H \rightarrow \lambda p q \cdot \mathfrak{S}(p, q, L)$. By induction with respect to m , $\mathfrak{S}(m, 1, I, I) \text{ conv } I$; using this fact, H will be found to have the desired properties.

such that $N(X) \vdash' \mathfrak{S}(S(X)) = \mathfrak{f}(X, \mathfrak{S}(X))$. Then $\mathfrak{S}(i, \lambda x \cdot F(x), m)$ conv $F(m) + F(m+1) + \cdots + F([m+i] - 1)$ ($m, i = 1, 2, \cdots$).

Let $\sum_{x=m}^n [R]$ be an abbreviation for $\mathfrak{S}([n+1] - m, \lambda x \cdot R, m)$, and define

$\prod_{x=m}^n [R]$ similarly, replacing the first occurrence of $+$ in \mathfrak{f} by \times .

16I. If m and n are positive integers and $m \leq n$, then $\sum_{x=m}^n F(x)$ conv $F(m) + F(m+1) + \cdots + F(n)$ and $\prod_{x=m}^n F(x)$ conv $F(m) \times F(m+1) \times \cdots \times F(n)$.

$$16.1: [N(\rho) \supset_{\rho} N(f(\rho))] \supset_f \cdot N(n) \supset_n \cdot N(\sum_{x=1}^n f(x)).$$

$$16.2: [N(\rho) \supset_{\rho} N(f(\rho))] \supset_f \cdot N(n) \supset_n \cdot \sum_{x=1}^{S(n)} f(x) = \sum_{x=1}^n f(x) + f(S(n)).$$

$$16.3: [N(\rho) \supset_{\rho} N(f(\rho))] \supset_f \cdot N(n) \supset_n \cdot S(\sum_{x=1}^n f(x)) > n.$$

Proofs. Assume $N(\rho) \supset_{\rho} N(f(\rho))$. Then (1) $N(f(1))$, and by conversion, $N(\sum_{x=1}^1 f(x))$. (2) Assume $N(n)$. Then $\sum_{x=1}^{S(n)} f(x)$ conv $\mathfrak{S}([S(n)+1] - 1, \lambda x \cdot f(x), 1) = \mathfrak{S}(S(n), \lambda x \cdot f(x), 1) = \mathfrak{f}(n, \mathfrak{S}(n), \lambda x \cdot f(x), 1)$, conv $\mathfrak{S}(n, \lambda x \cdot f(x), 1) + f(1+n)$, $= \mathfrak{S}([n+1] - 1, \lambda x \cdot f(x), 1) + f(S(n))$, which is $\sum_{x=1}^n f(x) + f(S(n))$. (3) Assuming $N(n)$ and $N(\sum_{x=1}^n f(x))$, and using (2) and 5.2, $N(\sum_{x=1}^{S(n)} f(x))$. (4) From (1) and (3), by induction, $N(n) \supset_n \cdot N(\sum_{x=1}^n f(x))$. Hence \vdash 16.1. (5) Assume $N(n)$. Using 16.1, $E(\sum_{x=1}^{S(n)} f(x))$. Hence, using (2) and § 2, $\sum_{x=1}^{S(n)} f(x) = \sum_{x=1}^n f(x) + f(S(n))$. Hence \vdash 16.2. (6) By (1) and 12.8, $S(\sum_{x=1}^1 f(x)) > 1$. Assume $N(n)$ and $S(\sum_{x=1}^n f(x)) > n$. Then $S^2(\sum_{x=1}^{S(n)} f(x)) = S^2(\sum_{x=1}^n f(x) + f(S(n)))$ (by (2)), $= S(\sum_{x=1}^n f(x)) + S(f(S(n)))$, $> S(\sum_{x=1}^n f(x)) + 1$ (12.8, 12.11), conv $S^2(\sum_{x=1}^n f(x))$, $> S(n)$ (by $S(\sum_{x=1}^n f(x)) > n$ and 12.11). Hence $S(\sum_{x=1}^{S(n)} f(x)) > S(n)$ (12.14). By induction, $N(n) \supset_n \cdot S(\sum_{x=1}^n f(x)) > n$.

$$16.4: N(k)[N(\rho) \supset_{\rho} \cdot f(\rho) = k] \supset_{fk} \cdot N(n) \supset_n \cdot \sum_{x=1}^n f(x) = nk.$$

Proof. Assume $N(k) \cdot N(p) \supset_p f(p) = k$. Then $\sum_{x=1}^1 f(x) \text{ conv } f(1)$,
 $= k, = 1k$ (6.1); and, assuming $N(n)$ and $\sum_{x=1}^n f(x) = nk$, $\sum_{x=1}^{S(n)} f(x) = \sum_{x=1}^n f(x)$
 $+ f(S(n))$ (16.2), $= nk + k, = S(n)k$. By induction, $N(n) \supset_n \sum_{x=1}^n f(x) = nk$.

17. Formal definition: successions of finite sequences. By 15III, we can find a \mathfrak{U} such that $\mathfrak{U}(1) \text{ conv } \lambda r p q m \cdot m(\delta_{S(q)}^{r(p)}, p, S(q))$ and $\mathfrak{U}(2) \text{ conv } \lambda r p q m \cdot I^q(m(\delta_1^{r(S(p))}, S(p), 1))$. Then, by 15IV, we can find a \mathfrak{B} such that $\mathfrak{B}(1) \text{ conv } \lambda r m \cdot m(\delta_1^{r(1)}, 1, 1)$, $\mathfrak{B}(k+1) \text{ conv } \lambda r \cdot \mathfrak{B}(k, r, \lambda \pi \cdot \mathfrak{U}(\pi, r))$ ($k = 2, 3, \dots$), and $N(\mathbf{X}) \vdash' \mathfrak{B}(S(\mathbf{X})) = \lambda r \cdot \mathfrak{B}(\mathbf{X}, r, \lambda \pi \cdot \mathfrak{U}(\pi, r))$. Let $\mathcal{Q} \rightarrow \lambda r n \cdot \mathfrak{B}(n, r, \lambda u v w \cdot I^u(f(v, w)))$.

17I. If \mathbf{R} defines the sequence r_1, r_2, \dots of positive integers, then $\mathcal{Q}(\mathbf{F}, \mathbf{R})$ defines the sequence $\mathbf{F}(1, 1), \mathbf{F}(1, 2), \dots, \mathbf{F}(1, r_1), \mathbf{F}(2, 1), \mathbf{F}(2, 2), \dots, \mathbf{F}(2, r_2), \dots$.

For, under the hypothesis, $\lambda n \cdot \mathfrak{B}(n, \mathbf{R})$ defines the sequence $\lambda m \cdot m(1, 1, 1)$, $\lambda m \cdot m(1, 1, 2), \dots, \lambda m \cdot m(1, 1, r_1 - 1)$, $\lambda m \cdot m(2, 1, r_1)$, $\lambda m \cdot m(1, 2, 1)$, $\lambda m \cdot m(1, 2, 2), \dots, \lambda m \cdot m(1, 2, r_2 - 1)$, $\lambda m \cdot m(2, 2, r_2), \dots$, from which fact the conclusion follows.

17.1: $[N(\xi) \supset_\xi N(r(\xi))] \supset_r N(p) \supset_p$
 $\cdot [x < S(p) \supset_x y < S(r(x)) \supset_y t(f(x, y))] \supset_{tt}$
 $\cdot z < S(\sum_{i=1}^p r(i)) \supset_z t(\mathcal{Q}(f, r, z)).$

Proof. Note that $N(\mathbf{l})N(\mathbf{p})N(\mathbf{q}) \vdash E(\lambda m \cdot m(\mathbf{l}, \mathbf{p}, \mathbf{q}))$. Assume $N(\xi) \supset_\xi N(r(\xi))$.

(-) Let $\mathfrak{C}_r \rightarrow \lambda p \sigma \cdot \mathfrak{B}([\sum_{i=1}^p r(i) + \min(\sigma, r(p))] - r(p), r) = \lambda m$
 $\cdot m(\delta_{\min(\sigma, r(p))}^{r(p)}, p, \min(\sigma, r(p)))$. (1a) $\mathfrak{B}([\sum_{i=1}^1 r(i) + \min(1, r(1))] - r(1), r) = \mathfrak{B}(1, r)$
 $\text{conv } \lambda m \cdot m(\delta_1^{r(1)}, 1, 1) = \lambda m \cdot m(\delta_{\min(1, r(1))}^{r(1)}, 1, \min(1, r(1)))$. Thus $\mathfrak{C}_r(1, 1)$. (b) Assume $N(\sigma)$ and $\mathfrak{C}_r(1, \sigma)$. Case 1: $\epsilon_{r(1)}^{S(\sigma)} = 2$. Then $S(\sigma) > r(1)$; consequently $\min(\sigma, r(1)) = r(1)$, $= \min(S(\sigma), r(1))$; and hence $\mathfrak{C}_r(1, S(\sigma))$ follows from $\mathfrak{C}_r(1, \sigma)$. Case 2: $\epsilon_{r(1)}^{S(\sigma)} = 1$. Then $S(r(1)) > S(\sigma), r(1) > \sigma, \min(S(\sigma), r(1)) = S(\sigma)$, and $\min(\sigma, r(1)) = \sigma$. Hence $\mathfrak{B}([\sum_{i=1}^1 r(i) + \min(S(\sigma), r(1))] - r(1), r) = \mathfrak{B}(\min(S(\sigma), r(1)), r)$,

$= \mathfrak{B}(S(\sigma), r), = \mathfrak{B}(S(\min(\sigma, r(1))), r), = \mathfrak{B}(\min(\sigma, r(1)), r, \lambda\pi \cdot \mathfrak{U}(\pi, r)).$
 (using the definition of \mathfrak{B}), $= \mathfrak{B}([\sum_{i=1}^1 r(i) + \min(\sigma, r(1))] - r(1), r, \lambda\pi \cdot \mathfrak{U}(\pi, r)),$
 $= \{\lambda m \cdot m(\delta_{\min(\sigma, r(1))}^{r(1)}, 1, \min(\sigma, r(1)))\}(\lambda\pi \cdot \mathfrak{U}(\pi, r))$ (by $\mathfrak{C}_r(1, \sigma)$),
 $= \{\lambda m \cdot m(\delta_{\sigma^{r(1)}}^{r(1)}, 1, \sigma)\}(\lambda\pi \cdot \mathfrak{U}(\pi, r)), \text{ conv } \mathfrak{U}(\delta_{\sigma^{r(1)}}^{r(1)}, r, 1, \sigma),$
 $= \mathfrak{U}(1, r, 1, \sigma)$ (since $\sigma < r(1)$), $\text{conv } \lambda m \cdot m(\delta_{S(\sigma)}^{r(1)}, 1, S(\sigma))$ (using the def. of \mathfrak{U}),
 $= \lambda m \cdot m(\delta_{\min(S(\sigma), r(1))}^{r(1)}, 1, \min(S(\sigma), r(1))).$ Thus $\mathfrak{C}_r(1, S(\sigma))$. Hence, by cases (C9I), $\mathfrak{C}_r(1, S(\sigma))$. (c) From (a) and (b) by induction, $N(\sigma) \supset_{\sigma} \mathfrak{C}_r(1, \sigma)$.
 (2) Assume $N(p)$ and $N(\sigma) \supset_{\sigma} \mathfrak{C}_r(p, \sigma)$. (a) $\mathfrak{B}([\sum_{i=1}^{S(p)} r(i) + \min(1, r(S(p)))] - r(S(p)), r) = \mathfrak{B}(S(\sum_{i=1}^n r(i)), r)$ (16.2, 11.2, 5.4, 13.2, 12.8),
 $= \mathfrak{B}(\sum_{i=1}^n r(i), r, \lambda\pi \cdot \mathfrak{U}(\pi, r))$ (by the def. of \mathfrak{B}), $= \mathfrak{B}([\sum_{i=1}^n r(i) + r(p)] - r(p), r, \lambda\pi \cdot \mathfrak{U}(\pi, r)),$
 $= \mathfrak{B}([\sum_{i=1}^n r(i) + \min(r(p), r(p))] - r(p), r, \lambda\pi \cdot \mathfrak{U}(\pi, r)),$
 $= \{\lambda m \cdot m(\delta_{\min(r(p), r(p))}^{r(p)}, p, \min(r(p), r(p)))\}(\lambda\pi \cdot \mathfrak{U}(\pi, r))$
 (by $N(\sigma) \supset_{\sigma} \mathfrak{C}_r(p, \sigma)$ and $N(r(p))$), $= \{\lambda m \cdot m(2, p, r(p))\}(\lambda\pi \cdot \mathfrak{U}(\pi, r)),$
 $\text{conv } \mathfrak{U}(2, r, p, r(p)), \text{ conv } \lambda m \cdot I^{r(p)}(m(\delta_1^{r(S(p))}, S(p), 1))$ (using the def. of \mathfrak{U}),
 $= \lambda m \cdot m(\delta_1^{r(S(p))}, S(p), 1), = \lambda m \cdot m(\delta_{\min(1, r(S(p)))}^{r(S(p))}, S(p), \min(1, r(S(p))))$. Thus $\mathfrak{C}_r(S(p), 1)$. (b) Assuming $N(\sigma)$ and $\mathfrak{C}_r(S(p), \sigma)$, $\mathfrak{C}_r(S(p), S(\sigma))$ follows by reasoning like that used in (1b) (in Case 2, 16.2 is used). (c) From (a) and (b) by induction, $N(\sigma) \supset_{\sigma} \mathfrak{C}_r(S(p), \sigma)$.
 (3) From (1) and (2) by induction, $N(p) \supset_p N(\sigma) \supset_{\sigma} \mathfrak{C}_r(p, \sigma)$. Thence we can infer $N(p) \supset_p \sigma < S(r(p)) \supset_{\sigma} \mathfrak{B}([\sum_{i=1}^p r(i) + \sigma] - r(p), r) = \lambda m \cdot m(\delta_{\sigma^{r(p)}}^{r(p)}, p, \sigma)$.

(ii) Let $\mathfrak{L}_{rz} \rightarrow \lambda p \cdot \sum ab \cdot a < S(p) \cdot b < S(r(a)) \cdot z = [\sum_{i=1}^a r(i) + b] - r(a)$. (a) Assume $z < S(\sum_{i=1}^1 r(i))$. Then $z = [r(1) + z] - r(1)$, $\text{conv } [\sum_{i=1}^1 r(i) + z] - r(1)$; also $1 < S(1)$ and $z < S(\sum_{i=1}^1 r(i))$, $\text{conv } S(r(1))$. Hence, using Axiom 14 and Rule IV, $\mathfrak{L}_{rz}(1)$. By Theorem I, $z < S(\sum_{i=1}^1 r(i)) \supset_z \mathfrak{L}_{rz}(1)$. (b) Assume $N(p)$, $z < S(\sum_{i=1}^p r(i)) \supset_z \mathfrak{L}_{rz}(p)$, and $z < S(\sum_{i=1}^{S(p)} r(i))$.
 Case 1: $\epsilon(S(\sum_{i=1}^2 r(i)), z) = 2$. Then $z < S(\sum_{i=1}^n r(i))$, and, using $z < S(\sum_{i=1}^n r(i)) \supset_z \mathfrak{L}_{rz}(p)$, we can prove $\mathfrak{L}_{rz}(S(p))$ by means of the second clause of Theorem I. Case 2: $\epsilon(S(\sum_{i=1}^n r(i)), z) = 1$. Then $z > \sum_{i=1}^n r(i)$, and

hence $z = \sum_{i=1}^n r(i) + \cdot z - \sum_{i=1}^n r(i)$ (12.5), $= [\sum_{i=1}^{S(p)} r(i) + \cdot z - \sum_{i=1}^n r(i)] - r(S(p))$ (16.2, 11.2, 5.4). Case A: $\epsilon(z - \sum_{i=1}^n r(i), r(S(p))) = 1$. Then $z - \sum_{i=1}^n r(i) < S(r(S(p)))$. Case B: $\epsilon(z - \sum_{i=1}^n r(i), r(S(p))) = 2$. Then $z - \sum_{i=1}^n r(i) > r(S(p))$. Hence $\sum_{i=1}^{S(p)} r(i) = \sum_{i=1}^n r(i) + r(S(p))$ (16.2), $< \sum_{i=1}^n r(i) + \cdot z - \sum_{i=1}^n r(i)$ (12.11), $= z$. Hence $\epsilon(z, \sum_{i=1}^{S(n)} r(i)) = 2$. But $\epsilon(z, \sum_{i=1}^{S(p)} r(i)) = 1$ is a consequence of the assumption $z < S(\sum_{i=1}^{S(p)} r(i))$. Hence, by cases A and B and *reductio ad absurdum* (C10II), $z - \sum_{i=1}^n r(i) < S(r(S(p)))$. Also $S(p) < S^2(p)$. Hence, using Axiom 14 and Rule IV, $\mathfrak{L}_{rz}(S(p))$. Hence, by cases 1 and 2 (C9I), $\mathfrak{L}_{rz}(S(p))$. By Thm. I, $z < S(\sum_{i=1}^{S(p)} r(i)) \supset \mathfrak{L}_{rz}(S(p))$. (c) From (a) and (b) by induction, $N(p) \supset_p \cdot z < S(\sum_{i=1}^p r(i)) \supset \mathfrak{L}_{rz}(p)$.

(iii) Assume $N(p)$, $x < S(p) \supset_x \cdot y < S(r(x)) \supset_y \cdot t(f(x, y))$, and $z < S(\sum_{i=1}^p r(i))$. Then, by (ii), $\sum ab \cdot a < S(p) \cdot b < S(r(a)) \cdot z = [\sum_{i=1}^a r(i) + b] - r(a)$. Assume $a < S(p) \cdot b < S(r(a)) \cdot z = [\sum_{i=1}^a r(i) + b] - r(a)$. Then $\mathcal{Q}(f, r, z) \text{ conv } \mathfrak{B}(z, r, \lambda uvw \cdot I^u(f(v, w))) = \mathfrak{B}([\sum_{i=1}^a r(i) + b] - r(a), r, \lambda uvw \cdot I^u(f(v, w))) = \{\lambda m \cdot m(\delta_b^{r(a)}, a, b)\}(\lambda uvw \cdot I^u(f(v, w)))$ (by (i)), $\text{conv } \delta_b^{r(a)}(I, f(a, b)) = f(a, b)$ (7.2). Moreover $t(f(a, b))$ is provable from our assumptions. Hence $t(\mathcal{Q}(f, r, z))$. By the second clause of Theorem I, $t(\mathcal{Q}(f, r, z))$ is provable without the last assumption.

$$17.2: [N(\xi) \supset_{\xi} N(r(\xi))] \supset_r \cdot [N(x) \supset_x \cdot y < S(r(x)) \supset_y \cdot t(f(x, y))] \supset_{ft} \cdot N(z) \supset_z \cdot t(\mathcal{Q}(f, r, z)).$$

Proof. Assuming $N(\xi) \supset_{\xi} N(r(\xi))$, $N(x) \supset_x \cdot y < S(r(x)) \supset_y \cdot t(f(x, y))$, and $N(z)$, we can prove $x < S(z) \supset_x \cdot y < S(r(x)) \supset_y \cdot t(f(x, y))$, and also, using 16.3, $z < S(\sum_{i=1}^z r(i))$. Hence, by 17.1, $t(\mathcal{Q}(f, r, z))$.

Using 17I, the dyads (triads, \dots) of positive integers can be *enumerated formally* (i. e., there is an enumeration of them which is definable formally). As another application of \mathcal{Q} , we establish the following theorem:

17II. If $A_1, \dots, A_l, R_1, \dots, R_{m+n}$ contain no free symbols, then a

formula H can be found such that (1) H enumerates formally (with repetitions) the formulas derivable from A_1, \dots, A_l by zero or more operations of passing from A and B to $R_1(A), \dots, R_m(A), R_{m+1}(A, B), \dots$, or $R_{m+n}(A, B)$, and (2) $T(A_1), \dots, T(A_l), T(a) \supset_a T(R_1(a)), \dots, T(a) \supset_a T(R_m(a)), T(a)T(b) \supset_{ab} T(R_{m+1}(a, b)), \dots, T(a)T(b) \supset_{ab} T(R_{m+n}(a, b)) \vdash N(z) \supset_z T(H(z))$.

Proof. Let $A_{1i} \rightarrow A_i$ ($i = 1, \dots, l_1$, where $l_1 = l$). Given A_{ki} ($i = 1, \dots, l_k$), let $A_{k+1,1}, \dots, A_{k+1,l_{k+1}}$, where $l_{k+1} = (1 + m + n)l_k^2$, be the formulas $A_{k1}, \dots, A_{kl_k}, \dots, A_{k1}, \dots, A_{kl_k}; R_1(A_{k1}), \dots, R_1(A_{kl_k}), \dots, R_1(A_{k1}), \dots, R_1(A_{kl_k}); \dots; R_m(A_{k1}), \dots, R_m(A_{kl_k}), \dots, R_m(A_{k1}), \dots, R_m(A_{kl_k}); R_{m+1}(A_{k1}, A_{k1}), \dots, R_{m+1}(A_{k1}, A_{kl_k}), \dots, R_{m+1}(A_{kl_k}, A_{k1}), \dots, R_{m+1}(A_{kl_k}, A_{kl_k}); \dots; R_{m+n}(A_{k1}, A_{k1}), \dots, R_{m+n}(A_{k1}, A_{kl_k}), \dots, R_{m+n}(A_{kl_k}, A_{k1}), \dots, R_{m+n}(A_{kl_k}, A_{kl_k})$ ($1 + m + n$ sets of l_k sets of l_k formulas each), respectively. Then the sequence of formulas A_{k1}, \dots, A_{kl_k} (defined by induction with respect to k) is an enumeration (with repetitions) of the formulas derivable from A_1, \dots, A_l by not more than $k - 1$ applications of the operations under consideration.

By 15III, there can be found a formula F_1 such that $F_1(i)$ conv A_{1i} ($i = 1, \dots, l_1$), and a formula J which defines the finite sequence $\lambda fji \cdot I^j(f(i)), \lambda fji \cdot R_1(I^j(f(i))), \dots, \lambda fji \cdot R_m(I^j(f(i))), \lambda fji \cdot R_{m+1}(f(j), f(i)), \dots, \lambda fji \cdot R_{m+n}(f(j), f(i))$. By 15IV, the sequence l_1, l_2, l_3, \dots can be defined by a formula L such that $N(X) \vdash' L(S(X)) = [1 + m + n]L(X)L(X)$. If $F_{k+1} \rightarrow \mathcal{Q}(\lambda v \cdot \mathcal{Q}(J(v, F_k), \lambda w \cdot I^w(L(k))), \lambda w \cdot I^w(L(k)L(k)))$ ($k = 1, 2, \dots$), then, by 15V, the sequence F_1, F_2, \dots is definable by a formula F such that $N(Y) \vdash' F(S(Y)) = \mathcal{Q}(\lambda v \cdot \mathcal{Q}(J(v, F(Y)), \lambda w \cdot I^w(L(Y))), \lambda w \cdot I^w(L(Y)L(Y)))$. Let $H \rightarrow \mathcal{Q}(F, L)$.

Assuming that $F_k(i)$ conv A_{ki} ($i = 1, \dots, l_k$), it follows by 17I and the definitions of F_{k+1} , J and L that $F_{k+1}(i)$ conv $A_{k+1,i}$ ($i = 1, \dots, l_{k+1}$). By induction with respect to k , $F_k(i)$ conv A_{ki} ($i = 1, \dots, l_k$; $k = 1, 2, \dots$). Hence, by 17I and the definitions of H , F and L , H defines $A_{11}, \dots, A_{1l_1}, A_{21}, \dots, A_{2l_2}, \dots$. Hence (1) is satisfied.

Assume $T(A_1), \dots, T(A_l), T(a) \supset_a T(R_1(a)), \dots, T(a) \supset_a T(R_m(a)), T(a)T(b) \supset_{ab} T(R_{m+1}(a, b)), \dots, T(a)T(b) \supset_{ab} T(R_{m+n}(a, b))$. In the following we suppose q, x and y to represent variables distinct from each other and from the variables of T . (1) $N(\xi) \supset_\xi N(L(\xi))$ can be proved by induction. (2) Using $T(A_1), \dots, T(A_l)$, we can prove $N(y) \supset_y T(F_1(\min(y, l)))$ by induction from an l -tuple basis, and thence infer $y < S(l) \supset_y T(F_1(y))$ by use of Theorem I and 13.2. By conversion, $y < S(L(1)) \supset_y \check{T}(F(1, y))$.

(3) Assume $N(q)$ and $y \leq S(L(q)) \supset_y T(F(q, y))$. (a) Assuming $x \leq S(L(q))$ and $y \leq S(L(q))$, we can infer $T(F(q, x))$ and $T(F(q, y))$; thence, using $T(a) \supset_a T(R_c(a)), T(R_c(F(q, y)))$ ($c = 1, \dots, m$), and, using $T(a)T(b) \supset_{ab} T(R_{m+d}(a, b)), T(R_{m+d}(F(q, x), F(q, y)))$ ($d = 1, \dots, n$); also, using the definition of $J, J(1, F(q), x, y) = F(q, y), J(1 + c, F(q), x, y) = R_c(F(q, y))$, and $J(1 + m + d, F(q), x, y) = R_{m+d}(F(q, x), F(q, y))$; hence $T(J(1, F(q), x, y)), T(J(1 + c, F(q), x, y)), T(J(1 + m + d, F(q), x, y))$. Thus, for $j = 1, \dots, 1 + m + n$, $T(J(j, F(q), x, y))$ is a consequence of $x \leq S(L(q)), y \leq S(L(q))$ and our other assumptions. Using these relations, we can prove by means of Thm. I and induction from a $1 + m + n$ -tuple basis, $N(v) \supset_v x \leq S(L(q)) \supset_x y \leq S(L(q)) \supset_y T(J(\min(v, 1 + m + n), F(q), x, y))$. (b) Assume $v \leq S(1 + m + n)$. Then from (a), by means of Theorem I, 13.2 and 7.2, we obtain $x \leq S(L(q)) \supset_x y \leq S(\{\lambda w \cdot I^w(L(q))\}(x)) \supset_y T(J(v, F(q), x, y))$. Using (1), $N(L(q))$; and hence, using 7.2 and Theorem I, $N(s) \supset_s N(\{\lambda w \cdot I^w(L(q))\}(s))$. These results with 17.1 yield $z \leq S(\sum_{u=1}^{L(q)} \{\lambda w \cdot I^w(L(q))\}(u)) \supset_z T(\mathcal{Q}(J(v, F(q)), \lambda w \cdot I^w(L(q)), z))$. Also, by using $N(L(q))$, 7.2 and Theorem I, $N(L(q)) \cdot N(s) \supset_s \{\lambda w \cdot I^w(L(q))\}(s) = L(q)$; hence $\sum_{u=1}^{L(q)} \{\lambda w \cdot I^w(L(q))\}(u) = L(q)L(q)$ (by 16.4), $= \{\lambda w \cdot I^w(L(q)L(q))\}(v)$. Using this result with the preceding, and applying Theorem I, $v \leq S(1 + m + n) \supset_v z \leq S(\{\lambda w \cdot I^w(L(q)L(q))\}(v)) \supset_z T(\{\lambda v \cdot \mathcal{Q}(J(v, F(q)), \lambda w \cdot I^w(L(q)))\}(v, z))$. (c) By Theorem I, $N(s) \supset_s N(\{\lambda w \cdot I^w(L(q)L(q))\}(s))$. Using the latter, $N(1 + m + n)$, and the result of (b) with 17.1, $z \leq S(\sum_{u=1}^{1+m+n} \{\lambda w \cdot I^w(L(q)L(q))\}(u)) \supset_z T(\mathcal{Q}(\lambda v \cdot \mathcal{Q}(J(v, F(q)), \lambda w \cdot I^w(L(q))), \lambda w \cdot I^w(L(q)L(q)), z))$. Thence, using the definition of F , Rule I, and the relation $\sum_{u=1}^{1+m+n} \{\lambda w \cdot I^w(L(q)L(q))\}(u) = [1 + m + n]L(q)L(q)$ (by 16.4), $= L(S(q))$ (by the def. of L), we infer $y \leq S(L(S(q))) \supset_y T(F(S(q), y))$. (4) From (2) and (3) by induction, $N(q) \supset_q y \leq S(L(q)) \supset_y T(F(q, y))$. This and (1) with 17.2 yield $N(z) \supset_z T(\mathcal{Q}(F, L, z))$, or, by the definition of H , $N(z) \supset_z T(H(z))$.

18. The sequence of positive integers satisfying a given condition. By 15III, there can be found a formula \mathfrak{F} such that

$$(1) \quad \begin{aligned} \mathfrak{F}(1) &\text{ conv } \lambda c d k \cdot c(1, d(k+1), c, d, k+1), \\ \mathfrak{F}(2) &\text{ conv } \lambda c d k \cdot c(2, I^{d(k)}, k), \end{aligned}$$

and then a formula \mathfrak{C} such that

$$(2) \quad \mathfrak{C}(1) \text{ conv } \mathfrak{J}, \quad \mathfrak{C}(2) \text{ conv } I.$$

Let $\mathfrak{p} \rightarrow \lambda dk : \mathfrak{J}(d(k), \mathfrak{C}, d, k)$.

18I. Given a positive integer k : If $D(k) \text{ conv } 2$, $\mathfrak{p}(D, k) \text{ conv } k$. If $D(k) \text{ conv } 1$, $\mathfrak{p}(D, k) \text{ conv } \mathfrak{p}(D, k+1)$. Hence, if $D(k) \text{ conv } D(k+1) \text{ conv } \dots \text{ conv } D(l-1) \text{ conv } 1$ and $D(l) \text{ conv } 2$ ($l \geq k$), then $\mathfrak{p}(D, k) \text{ conv } l$.

For if $D(k) \text{ conv } 2$, then $\mathfrak{p}(D, k) \text{ conv } \mathfrak{J}(D(k), \mathfrak{C}, D, k) \text{ conv } \mathfrak{J}(2, \mathfrak{C}, D, k) \text{ conv } \mathfrak{C}(2, I^{D(k)}, k) \text{ conv } I(I^2, k) \text{ conv } k$; and if $D(k) \text{ conv } 1$, then $\mathfrak{p}(D, k) \text{ conv } \mathfrak{J}(1, \mathfrak{C}, D, k) \text{ conv } \mathfrak{C}(1, D(k+1), \mathfrak{C}, D, k+1) \text{ conv } \mathfrak{J}(D(k+1), \mathfrak{C}, D, k+1) \text{ conv } \mathfrak{p}(D, k+1)$.

18II. If $D(i) \text{ conv } 1$ for every positive integer $i \geq$ the positive integer k , then $\mathfrak{p}(D, k)$ has no normal form.*

Proof. A derivation of B from A by applications of I and II, including at least one of the latter, will be called a *reduction*. A conversion in which III is not used may be indicated by an accent. It will be shown in a forthcoming paper by A. Church and J. B. Rosser,† that if an expression A has a normal form, then every sequence $A \text{ red } A' \text{ red } A'' \text{ red } \dots$ of reductions is finite;‡ and that if $P \text{ conv } Q$, then there exists a conversion of P into Q in which all applications of III follow all applications of II.§ Hence if \bar{A} is a normal form of A , $A \text{ conv}' \bar{A}$. $\lambda cdk : c(1, d(k+1), c, d, k+1)$ is a normal form of $\mathfrak{J}(1)$. Consequently \mathfrak{J} has a normal form $\bar{\mathfrak{J}}$, for otherwise there would exist an infinite sequence $\mathfrak{J} \text{ red } \mathfrak{J}' \text{ red } \mathfrak{J}'' \text{ red } \dots$, and hence an infinite sequence $\mathfrak{J}(1) \text{ red } \mathfrak{J}'(1) \text{ red } \mathfrak{J}''(1) \text{ red } \dots$. (1) and (2) hold with \mathfrak{J} replaced by $\bar{\mathfrak{J}}$, and *conv* by *conv'*. Moreover $\bar{i} + 1 \text{ conv}' \bar{i} + 1$, and from $D(i) \text{ conv } 1$ follows $D(\bar{i}) \text{ conv}' 1$. Then under the hypothesis, $\mathfrak{p}(D, i) \text{ red } \mathfrak{J}(D(i), \mathfrak{C}, D, i) \text{ conv}' \bar{\mathfrak{J}}(D(\bar{i}), \mathfrak{C}, D, \bar{i}) \text{ conv}' \bar{\mathfrak{J}}(1, \mathfrak{C}, D, \bar{i}) \text{ red } \mathfrak{C}(1, D(\bar{i}+1), \mathfrak{C}, D, \bar{i}+1) \text{ conv}' \bar{\mathfrak{J}}(D(\bar{i}+1), \mathfrak{C}, D, \bar{i}+1) \text{ conv}' \bar{\mathfrak{J}}(D(\bar{i}+1), \mathfrak{C}, D, \bar{i}+1)$. Hence $\mathfrak{p}(D, k) \text{ red } \bar{\mathfrak{J}}(D(\bar{k}), \mathfrak{C}, D, \bar{k}) \text{ red } \bar{\mathfrak{J}}(D(\bar{k}+1), \mathfrak{C}, D, \bar{k}+1) \text{ red } \bar{\mathfrak{J}}(D(\bar{k}+2), \mathfrak{C}, D, \bar{k}+2) \text{ red } \dots \text{ ad infinitum}$, which could not be if $\mathfrak{p}(D, k)$ had a normal form.

* Normal form is defined in § C5.

† A. Church and J. B. Rosser, "Some properties of conversion."

‡ In other words, given any well-formed expression P , either all or none of the sequences $P \text{ red } P' \text{ red } P'' \text{ red } \dots$ can be continued *ad infinitum*.

§ Consequently, if A has a normal form, all normal forms of A are derivable from a given one by applications of I.

By 15IV, a formula \mathfrak{A} such that $\mathfrak{A}(1) \text{ conv } \lambda d \cdot \mathfrak{p}(d, 1)$ and $\mathfrak{A}(n+1) \text{ conv } \lambda d \cdot \mathfrak{p}(d, \mathfrak{A}(n, d) + 1)$ ($n = 1, 2, \dots$) can be found. Let $\mathcal{P} \rightarrow \lambda d n \cdot \mathfrak{A}(n, d)$.

18III. If \mathbf{D} defines the infinite sequence d_1, d_2, d_3, \dots of 1's and 2's, and $d_{n_1}, d_{n_2}, d_{n_3}, \dots$ is the subsequence which are 2's, then $\mathcal{P}(\mathbf{D})$ defines the sequence n_1, n_2, n_3, \dots . If the latter is a finite sequence n_1, \dots, n_k ($k \geq 0$), then, for $i > k$, $\mathcal{P}(\mathbf{D}, i)$ has no normal form.

This result, together with 15Ij, l and above results concerning the formal definability and enumerability of n -tuples of positive integers, leads to the following:

18IV. Given functions $F_i(x_1, \dots, x_n)$ and $G_i(x_1, \dots, x_n)$ ($i = 1, \dots, m$) which are defined for all n -tuples of positive integers (x_1, \dots, x_n) and whose values are positive integers, if F_i and G_i are definable formally, then there can be found a formula \mathbf{L} such that (a) if solutions of the system of equations

$$(3) \quad F_i(x_1, \dots, x_n) = G_i(x_1, \dots, x_n) \quad (i = 1, \dots, m)$$

exist, \mathbf{L} enumerates them formally,* and (b) if less than k different solutions exist, $\mathbf{L}(k)$ does not have a normal form.

For example, a formula \mathfrak{F} can be found such that (a) \mathfrak{F} enumerates the solutions of $x^t + y^t = z^t$ ($t \geq 2$) in positive integers, if such solutions exist, and (b) the Fermat problem is equivalent to the problem of whether $\mathfrak{F}(1)$ has a normal form.

We have noted that a theory of formal definition of functions of natural numbers, similar to our theory for functions of positive integers, can be constructed. It is also easy to construct a like theory for integers, if the integer x is represented by the formula $[\mathbf{x}_1, \mathbf{x}_2]$, where x_1, x_2 are the least positive integers such that $x_1 - x_2 = x$; and a like theory for rational numbers, if the rational number x is represented by the formula $[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$ where x_1, x_2, x_3 are the least positive integers such that $(x_1 - x_2)/x_3 = x$. In particular, theorems corresponding to 18IV can be proved for each of these theories.

Given any formula T in the notation of *Principia Mathematica*, there can be found a well-formed expression \mathbf{K} such that the problem whether T is provable in the system of *Principia* is equivalent to the problem, whether \mathbf{K} has a normal form. Indeed, suppose we have given any formula T and any system of formal logic F , for which the condition is satisfied that there is a

* That is, there is an enumeration of the solutions as $(x_{j_1}, \dots, x_{j_n})$ ($j = 1, 2, \dots$) such that $L(j) \text{ conv } [\mathbf{x}_{j_1}, \dots, \mathbf{x}_{j_n}]$, in the notation of § 8.

class M of formulas such that (a) all provable formulas of F belong to M , (b) T belongs to M , and (c) there exists a one-to-one correspondence of M to a class of positive integers such that the numbers corresponding to provable formulas are enumerable formally in the sense of the present theory (let t correspond to T , and L enumerate the numbers ordered to provable formulas). Then the problem whether T is provable in F is equivalent to the problem whether $\mathcal{P}(\lambda n \cdot \delta_t^{L(n)}, 1)$ has a normal form.

19. A representation of the logic C_1 within itself. Let C_1 denote the logic whose formal axioms are 1, 3-11, 14-16, and whose rules of procedure are I-V.

The objective of this section is its last theorem, to establish which we utilize a representation of the logic C_1 within itself in the fashion of Gödel.* Our particular choice of a representation serves to simplify the formal proofs. Instead of setting it up directly, we first set up a representation of the combinations without free symbols by formulas which will be called "metads," and then avail ourselves of a relation suggested by Rosser between C_1 and a certain system of combinations without free symbols.

Let r be an expression such that $r(1) \text{ conv } \lambda m \cdot m(\lambda pq \cdot I^q(p))$ and $r(S(k)) \text{ conv } \lambda m \cdot m(\lambda pqr \cdot Ir^{(k,q)}(Ir^{(k,r)}(p)))$, and h an expression such that $h(1) \text{ conv } \lambda p \cdot r(1, \lambda m \cdot m(1, p))$ and $h(S(k)) \text{ conv } \lambda pq \cdot r(S(k), \lambda m \cdot m(S(k), p, q))$ ($k = 1, 2, 3, \dots$).† Abbreviate $a(h)$ to $|a|$, $\{\lambda xm \cdot m(1, x)\}(x)$ to $[x]$, $\{\lambda abm \cdot m(S(|a|), a, b)\}(a, b)$ to $[a, b]$, and $[[x_1, \dots, x_{2^{r-1}}], [x_{2^{r-1}+1}, \dots, x_{2^r}]]$ to $[x_1, \dots, x_{2^r}]$. A formula a shall be called a *metad* (of rank r) if $a \text{ conv } [x_1, \dots, x_{2^r}]$, where $x_1, \dots, x_{2^{r-1}}$ is a set of 1's and 2's.

19I. If a is a metad of rank r , then $|a| \text{ conv } r$.

For, by induction with respect to r , if a is a metad of rank r , then $r(r, a) \text{ conv } r$ and $|a| \text{ conv } r(r, a)$ (cf. the proof of 19.1).

Let $\text{ad} \rightarrow \lambda \mu \cdot [\phi([1])\phi([2])] \cdot [\phi(a)\phi(b) | a| = |b|] \supset_{ab} \phi([a, b]) \supset \phi \phi(\mu)$.

19.1(x): $\text{ad}([x]) \quad (x = 1, 2).$

* *Loc. cit.*

† Henceforth the introduction of expressions in accordance with the Theorems 15III-15V will be made in an abbreviated manner, as here where r is supposed to satisfy not only the stated relations but also the relation $N(K) \vdash' r(S(K)) = \lambda m \cdot m(\lambda pqr \cdot r(K, q, I, r(K, r, I, p)))$ (cf. 15IV), and h is supposed to satisfy not only the stated relations but also the relation $N(K) \vdash' h(S(K)) = \lambda pq \cdot r(S(K), \lambda m \cdot m(S(K), p, q))$ (cf. 15III).

19.2: $\text{ad}(a) \supset_a N(|a|).$

Proofs. (1) $N(|[x]|) \cdot |[x]| = \mathbf{r}(|[x]|, [x])$ ($x = 1, 2$) is provable by conversion from $N(1) \cdot 1 = 1$. (2) Assume $N(|a|)$, $|a| = \mathbf{r}(|a|, a)$, $N(|b|)$, $|b| = \mathbf{r}(|b|, b)$, $|a| = |b|$. Then $[a, b] \text{ conv } \mathfrak{h}(S(|a|), a, b)$, $= \{\lambda p q \cdot \mathbf{r}(S(|a|), \lambda m \cdot m(S(|a|), p, q))\}(a, b)$ (using the assumption $N(|a|)$ and the last property of \mathfrak{h} as selected in accordance with 15III), $\text{conv } \mathbf{r}(S(|a|), [a, b])$, $= \{\lambda m \cdot m(\lambda p q r \cdot I^{\mathbf{r}(|a|, q)}(I^{\mathbf{r}(|a|, r)}(p)))\}([a, b])$ (using $N(|a|)$ and the last property of \mathbf{r} as selected in accordance with 15IV), $\text{conv } I^{\mathbf{r}(|a|, a)}(I^{\mathbf{r}(|a|, b)}(S(|a|)))$, $= I^{\mathbf{r}(|a|, a)}(I^{\mathbf{r}(|b|, b)}(S(|a|)))$ (by the assumption $|a| = |b|$), $= I^{|a|}(I^{|b|}(S(|a|)))$ (by the assumptions $|a| = \mathbf{r}(|a|, a)$ and $|b| = \mathbf{r}(|b|, b)$), $= I(I(S(|a|)))(N(|a|), N(|b|), 7.2)$, $\text{conv } S(|a|)$. Hence $[a, b] = \mathbf{r}([a, b], [a, b])$ (note the occurrence of $\mathbf{r}(S(|a|), [a, b])$ in the foregoing chain of equalities) and $N(|[a, b]|)$ (using $N(|a|)$ and 3.2). (3) Hence, if $\mathfrak{D}_\phi \rightarrow \phi([1])\phi([2]) \cdot [\phi(a)\phi(b) |a| = |b|] \supset_{ab} \phi([a, b])$, $\{\lambda \phi \cdot \mathfrak{D}_\phi\}(\lambda \alpha \cdot N(|\alpha|) \cdot |\alpha| = \mathbf{r}(|\alpha|, \alpha))$ is provable. Hence $\Sigma \phi \cdot \mathfrak{D}_\phi$. $\mathfrak{D}_\phi \vdash \phi([x])$ ($x = 1, 2$). By Theorem I, $\text{ad}([x])$. (4) Now $\Sigma a \cdot \text{ad}(a)$. Assume $\text{ad}(a)$. From $\text{ad}(a)$ and $\{\lambda \phi \cdot \mathfrak{D}_\phi\}(\lambda \alpha \cdot N(|\alpha|) \cdot |\alpha| = \mathbf{r}(|\alpha|, \alpha))$ by Rule V, $\{\lambda \phi \cdot \phi(a)\}(\lambda \alpha \cdot N(|\alpha|) \cdot |\alpha| = \mathbf{r}(|\alpha|, \alpha))$. Thence, $N(|a|)$.

19.3: $[\text{ad}(a)\text{ad}(b) |a| = |b|] \supset_{ab} \text{ad}([a, b]).$

19.4: $[\phi([1])\phi([2])$
 $\cdot [\text{ad}(a)\phi(a)\text{ad}(b)\phi(b) |a| = |b|] \supset_{ab} \phi([a, b])] \supset_\phi \cdot \text{ad}(c) \supset_c \phi(c).$

These theorems follow from 19.1 and the formula $\Sigma \phi \cdot \mathfrak{D}_\phi$ occurring in the proof of 19.1 in the same manner as 3.2 and 3.3 from 3.1 and \mathfrak{A}_5 of the proof of 3.1. The inference of an expression of the form $\text{ad}(c) \supset_c F(c)$ by means of 19.4 will be said to be by *induction (with respect to c)*.

Choose \mathfrak{m}_j so that $\mathfrak{m}_j(1) \text{ conv } I$, $\mathfrak{m}_1(S(\mathbf{k})) \text{ conv } \lambda m \cdot m(\lambda p q r \cdot I^p(I^{\mathbf{r}}(q)))$, $\mathfrak{m}_2(S(\mathbf{k})) \text{ conv } \lambda m \cdot m(\lambda p q r \cdot I^p(I^{\mathbf{r}}(r)))$, and let $\mathfrak{M}_j \rightarrow \lambda \rho \cdot \mathfrak{m}_j(|\rho|, \rho)$ ($j = 1, 2$; $k = 1, 2, 3, \dots$)*. We abbreviate $\mathfrak{M}_j(a)$ to \mathbf{a}_j , $\mathfrak{M}_i(\mathfrak{M}_j(a))$ to \mathbf{a}_{ji} , etc., when \mathbf{a} is a metad or represents a metad in the formal argument.

19II. If $\mathbf{a} \text{ conv } [x_1, \dots, x_{2^{r-1}}]$, where $x_1, \dots, x_{2^{r-1}}$ are 1's and 2's and $r > 1$, then $\mathbf{a}_1 \text{ conv } [x_1, \dots, x_{2^{r-2}}]$ and $\mathbf{a}_2 \text{ conv } [x_{2^{r-2}+1}, \dots, x_{2^{r-1}}]$.

19.5: $[\text{ad}(a)\text{ad}(b) |a| = |b|] \supset_{ab} |[a, b]| = S(|a|) \cdot [a, b]_1 = a \cdot [a, b]_2 = b.$

* We know that there exists an expression \mathfrak{m}_1 having the specified properties, and the property $N(\mathbf{K}) \vdash \mathfrak{m}_1(S(\mathbf{K})) = \lambda m \cdot m(\lambda p q r \cdot I^p(I^{\mathbf{r}}(q)))$, by use of 15III in conjunction with 15Ie and 7.2, taking for F the expression $\lambda x \cdot I^x(\lambda m \cdot m(\lambda p q r \cdot I^p(I^{\mathbf{r}}(q))))$. The introduction of \mathfrak{m}_2 is justified in the same manner.

Proof. Assume $\text{ad}(a) \text{ ad}(b) | a | = | b |$. By (4) of the proof of 19.1 and 19.2, we can infer $N(| a |), | a | = \mathbf{r}(| a |, a), N(| b |), | b | = \mathbf{r}(| b |, b)$, and hence, by (2) of the same proof, $|[a, b]| = S(| a |)$. Then also $[a, b]_1 \text{ conv } \mathbf{m}_1(|[a, b]|, [a, b]), = \mathbf{m}_1(S(| a |), [a, b]), = \{\lambda m \cdot m(\lambda pqr \cdot I^p(I^r(q)))\}([a, b])$ (by a supposition concerning \mathbf{m}_1), $\text{conv } I^{S(| a |)}(I^{| b |}(a))$, $= I(I(a))$ (19.2, 7.2), $\text{conv } a. \text{ ad}(a) \vdash E(a)$. Hence, by § 2, $[a, b]_1 = a$. Similarly $[a, b]_2 = b$.

Let $\mathbf{c} \rightarrow \lambda m \cdot m(\lambda p \cdot I^p)$.

19.6: $[\text{ad}(a) | a | = 1] \supset_a \cdot \mathbf{c}(a) < 3 \cdot a = [\mathbf{c}(a)]$.

19.7: $[\text{ad}(a) | a | > 1] \supset_a \cdot a = [a_1, a_2] \cdot \text{ad}(a_1) \cdot \text{ad}(a_2) \cdot | a_1 | = | a_2 | = | a | - 1$.

Proofs. Choose an expression \mathfrak{B} such that $\mathfrak{B}(1) \text{ conv } \lambda a \cdot \mathbf{c}(a) < 3 \cdot a = [\mathbf{c}(a)] \cdot E(I)$ and $\mathfrak{B}(2) \text{ conv } \lambda a \cdot a = [a_1, a_2] \cdot \text{ad}(a_1) \cdot \text{ad}(a_2) \cdot | a_1 | = | a_2 | = | a | - 1$. Then, using 19.5, the lemma $\text{ad}(a) \supset_a \mathfrak{B}(\epsilon_1 | a |, a)$ can be proved by induction. 19.6 and 19.7 follow.

19.8: $\text{ad}(a) \supset_a \cdot \text{ad}(a_1) \text{ ad}(a_2)$.

Proof. Assume $\text{ad}(a)$. Case 1: $\epsilon_1 | a | = 1$. Then $| a | = 1$; and, using the definitions of \mathfrak{M}_j and \mathbf{m}_j , $a = a_j$ ($j = 1, 2$). Using the latter, $\text{ad}(a_j)$ follows from $\text{ad}(a)$. Case 2: $\epsilon_1 | a | = 2$. Then $| a | > 1$; and $\text{ad}(a_1) \text{ ad}(a_2)$ can be proved by means of 19.7.

In the remainder of this paper, we shall mean by a *combination* a combination, in the sense of § C6, which contains no free symbols; in other words, a combination whose terms are I 's and J 's. If \mathbf{T} is the only term of a combination, the *rank* of \mathbf{T} shall be 1; if \mathbf{T} is a term of \mathbf{M} of rank r , then the rank of \mathbf{T} as a term of $\{\mathbf{M}\}(\mathbf{N})$ or $\{\mathbf{N}\}(\mathbf{M})$ shall be $r + 1$. The *rank* of a combination shall be the rank of its term of highest rank. A combination shall be *uniform* if all its terms have the same rank. (A uniform combination \mathbf{A}' of rank r has 2^{r-1} terms, and they occur in \mathbf{A}' in a linear series—cf. C6III). A uniform combination \mathbf{A}' shall *represent* a combination \mathbf{A} , if \mathbf{A}' is derivable from \mathbf{A} by zero or more substitutions of $I(\mathbf{T})$ for \mathbf{T} , where \mathbf{T} is a term. Given the correspondence $\begin{pmatrix} I & J \\ 1 & 2 \end{pmatrix}$, $[x_1, \dots, x_{2^{r-1}}]$ shall *correspond* to a uniform combination \mathbf{A}' if $x_1, \dots, x_{2^{r-1}}$ is the series of the numbers which correspond to the respective terms of \mathbf{A}' . If \mathbf{A} is a combination and $[x_1, \dots, x_{2^{r-1}}]$ corresponds to a uniform combination \mathbf{A}' which represents \mathbf{A} , we write " $[x_1, \dots, x_{2^{r-1}}] \sim \mathbf{A}$." A metad \mathbf{a} shall *represent* a combination \mathbf{A} if $\mathbf{a} \text{ conv } [x_1, \dots, x_{2^{r-1}}]$ where $[x_1, \dots, x_{2^{r-1}}] \sim \mathbf{A}$.

19III. Suppose that x_i, y_i ($i = 1, \dots, 2^{r-1}$) are 1's and 2's. a. Given a combination A , a representing metad a can be found. b. If the metad a represents the combination A , then a is of rank \geq the rank of A . c. If $[x_1, \dots, x_{2^{r-1}}] \sim A$ and $[y_1, \dots, y_{2^{r-1}}] \sim B$, then $[x_1, \dots, x_{2^{r-1}}, y_1, \dots, y_{2^{r-1}}] \sim \{A\}(B)$. d. If $[x_1, x_2, \dots, x_{2^{r-1}}] \sim A$, then $[1, x_1, 1, x_2, \dots, 1, x_{2^{r-1}}] \sim A$. e. If both $[x_1, \dots, x_{2^{r-1}}] \sim A$ and $[y_1, \dots, y_{2^{r-1}}] \sim A$, then $x_i = y_i$. f. If the metad a represents the combination $\{F\}(P)$, then a_1 represents F , and a_2 represents P .

Let c be an expression such that $c(1) \text{ conv } \lambda p \cdot [[1], p]$ and $c(S(k)) \text{ conv } \lambda p \cdot [c(k, p_1), c(k, p_2)]$ ($k = 1, 2, \dots$). Let $\mathfrak{G} \rightarrow \lambda p \cdot c(|p|, p)$.

19IV. If $a \text{ conv } [x_1, x_2, \dots, x_{2^{r-1}}]$ ($x_1, \dots, x_{2^{r-1}}$ being 1's and 2's), then $\mathfrak{G}(a) \text{ conv } [1, x_1, 1, x_2, \dots, 1, x_{2^{r-1}}]$.

The proof is by induction with respect to r (using 19I and 19II).

19.9: $N(\rho) \supset_p \text{ad}(a) \supset_a \text{ad}(\mathfrak{G}(a)) \cdot |\mathfrak{G}(a)| = \rho + |a|$.

Proof. $\text{ad}(a) \supset_a \text{ad}(\mathfrak{G}(a)) \cdot |\mathfrak{G}(a)| = S(|a|)$ is provable by induction with respect to a (using 19.1-19.3, 19.5), and 19.9 follows by induction with respect to ρ .

19.10: $[N(\rho) \text{ad}(a) \text{ad}(b) \cdot \mathfrak{G}(a) = \mathfrak{G}(b)] \supset_{\rho ab} a = b$.

Proof. Let c' be an expression such that $c'(1) \text{ conv } \lambda p \cdot p_2$ and $c'(S(k)) \text{ conv } \lambda p \cdot [c'(k, p_1), c'(k, p_2)]$ ($k = 1, 2, \dots$). Let $\mathfrak{G}' \rightarrow \lambda p \cdot c'(|p| - 1, p)$. Then $\text{ad}(a) \supset_a \mathfrak{G}'(\mathfrak{G}(a)) = a$ is provable by induction with respect to a , and $N(\rho) \supset_p \text{ad}(a) \supset_a \mathfrak{G}'(\mathfrak{G}(a)) = a$ follows by induction with respect to ρ . 19.10 follows from the latter in the same manner as 11.4 from 11.2.

Let $\langle a, b \rangle \rightarrow [\mathfrak{G}^{|b|}(a), \mathfrak{G}^{|a|}(b)]$.

19V. If the metads a and b represent the combinations A and B , respectively, then $\langle a, b \rangle$ is a metad which represents $\{A\}(B)$.

This follows from 19I, 15Id, 19IV, 19IIId, c.

19.11: $\text{ad}(a) \text{ad}(b) \supset_{ab} \text{ad}(\langle a, b \rangle)$.

Let \mathfrak{D} be an expression such that $\mathfrak{D}(1) \text{ conv } \lambda pq \cdot \delta(p(\lambda n \cdot I^n), q(\lambda n \cdot I^n))$ and $\mathfrak{D}(S(k)) \text{ conv } \lambda pq \cdot \mathfrak{D}(k, p_1, q_1) \circ \mathfrak{D}(k, p_2, q_2)$ ($k = 1, 2, \dots$). Let $\Delta \rightarrow \lambda pq \cdot \mathfrak{D}(|p| + |q|, \mathfrak{G}^{|q|}(p), \mathfrak{G}^{|p|}(q))$, and abbreviate $\Delta(a, b)$ to Δ_a^b .

• 19VI. If the metads a and b both represent the combination A , then $\Delta_a^a \text{ conv } 2$.

Proof. By induction with respect to r (using 15Ie, 19II), if $x_1, \dots, y_{2^{r-1}}$ are 1's and 2's, $\mathfrak{D}(r, [x_1, \dots, x_{2^{r-1}}], [y_1, \dots, y_{2^{r-1}}]) \text{ conv } \delta_{y_1}^{x_1} \circ \dots \circ \delta_{y_{2^{r-1}}}^{x_{2^{r-1}}}$. Hence, by 15II, j and 19IIIe, if $[x_1, \dots, x_{2^{r-1}}] \sim A$ and $[y_1, \dots, y_{2^{r-1}}] \sim A$, then $\mathfrak{D}(r, [x_1, \dots, x_{2^{r-1}}], [y_1, \dots, y_{2^{r-1}}]) \text{ conv } 2$. Moreover, by 19IV, 19I, 15Id and 19IIId, if $a \text{ conv } [x'_1, \dots, x'_{2^{m-1}}], [x'_1, \dots, x'_{2^{m-1}}] \sim A$, $b \text{ conv } [y'_1, \dots, y'_{2^{n-1}}], [y'_1, \dots, y'_{2^{n-1}}] \sim A$, then there are $x_1, \dots, x_{2^{r-1}}, y_1, \dots, y_{2^{r-1}}$ ($r = m + n$) such that $\mathfrak{G}^{[b]}(a) \text{ conv } [x_1, \dots, x_{2^{r-1}}], [x_1, \dots, x_{2^{r-1}}] \sim A$, $\mathfrak{G}^{[a]}(b) \text{ conv } [y_1, \dots, y_{2^{r-1}}], [y_1, \dots, y_{2^{r-1}}] \sim A$.

$$19.12: \quad \text{ad}(a)\text{ad}(b) \supset_{ab} \cdot M(\Delta_b^a).$$

$$19.13: \quad \text{ad}(a)\text{ad}(b) \supset_{ab} \cdot \Delta_b^a = \Delta_a^b.$$

$$19.14: \quad \text{ad}(a)\text{ad}(b) \supset_{ab} \cdot N(\rho) \supset_{\rho} \cdot \Delta(a, \mathfrak{G}^{\rho}(b)) = \Delta_b^a.$$

Proofs. If $\mathfrak{B} \rightarrow \lambda a \cdot [\text{ad}(b) | a | = | b |] \supset_b \cdot M(\mathfrak{D}(| a |, a, b)) \cdot \mathfrak{D}(| a |, a, b) = \mathfrak{D}(| a |, b, a) \cdot \mathfrak{D}(S(| a |), \mathfrak{G}(a), \mathfrak{G}(b)) = \mathfrak{D}(| a |, a, b)$, the lemma $\text{ad}(a) \supset_a \mathfrak{B}(a)$ can be proved by induction, using first 19.6, 14.10, 14.11, 14.5, and then the relation $\text{ad}(l), \text{ad}(m), | l | = | m |, \text{ad}(b), |[l, m]| = | b | \vdash \mathfrak{D}([l, m], [l, m], b) = \mathfrak{D}(| l |, l, b_1) \circ \mathfrak{D}(| m |, m, b_2) \cdot \mathfrak{D}([l, m], b, [l, m]) = \mathfrak{D}(| l |, b_1, l) \circ \mathfrak{D}(| m |, b_2, m) \cdot \mathfrak{D}(S([l, m]), \mathfrak{G}([l, m]), \mathfrak{G}(b)) = \mathfrak{D}(S(| l |), \mathfrak{G}(l), \mathfrak{G}(b_1)) \circ \mathfrak{D}(S(| m |), \mathfrak{G}(m), \mathfrak{G}(b_2))$ (which follows from 19.2, 19.5, 19.9, 19.7), and 19.2, 19.3, 19.5, 19.7, 14.2. 19.12-19.14 follow from the lemma, 19.2, 19.9, and the relation $\text{ad}(a)\text{ad}(b) \vdash \Delta(a, \mathfrak{G}(b)) = \mathfrak{D}(S(| \mathfrak{G}^{[b]}(a) |), \mathfrak{G}(\mathfrak{G}^{[b]}(a)), \mathfrak{G}(\mathfrak{G}^{[a]}(b)))$.

$$19.15: \quad [\text{ad}(a)\text{ad}(b) \cdot | a | = | b | \cdot \Delta_b^a = 2] \supset_{ab} \cdot a = b.$$

$$19.16: \quad \text{ad}(a) \supset_a \cdot \Delta_a^a = 2.$$

Proofs. If $\mathfrak{C} \rightarrow \lambda a \cdot [\mathfrak{D}(| a |, a, a) = 2] \cdot [\text{ad}(b) \cdot | a | = | b | \cdot \mathfrak{D}(| a |, a, b) = 2] \supset_b \cdot a = b$, then $\text{ad}(a) \supset_a \mathfrak{C}(a)$ is provable by induction, using first 19.6, 14.14, and then $\text{ad}(a) \supset_a \mathfrak{B}(a)$, 19.2, 19.3, 19.5, 19.7, 14.6 and the relation $\text{ad}(l), \text{ad}(m), | l | = | m |, \text{ad}(b), |[l, m]| = | b | \vdash \mathfrak{D}([l, m], [l, m], b) = \mathfrak{D}(| l |, l, b_1) \circ \mathfrak{D}(| m |, m, b_2)$. 19.15 and 19.16 follow, using 19.10.

Let \mathfrak{F} be an expression such that $\mathfrak{F}(1) \text{ conv } I$ and $\mathfrak{F}(2) \text{ conv } J$, and \mathfrak{g} an expression such that $\mathfrak{g}(1) \text{ conv } \lambda a \cdot a(\lambda p q \cdot I^p(\mathfrak{F}(q)))$ and $\mathfrak{g}(S(k)) \text{ conv } \lambda a \cdot \mathfrak{g}(k, a_1, \mathfrak{g}(k, a_2))$ ($k = 1, 2, \dots$). Let $\mathfrak{G} \rightarrow \lambda a \cdot \mathfrak{g}(| a |, a)$.

19VII. If the metad a represents the combination A , then $\mathfrak{G}(a) \text{ conv } A$.

For, by induction with respect to r , if $[x_1, \dots, x_{2^{r-1}}]$ corresponds to a uniform combination A' , $\mathfrak{G}([x_1, \dots, x_{2^{r-1}}]) \text{ conv } A'$. If A' represents A , $A' \text{ conv } A$.

Let i be an expression such that $i(1) \text{ conv } [1]$ and $i(S(k)) \text{ conv } [i(k), i(k)]$ ($k = 1, 2, \dots$).

$$19.17: \quad N(r) \supset_r \cdot \text{ad}(i(r)) \cdot |i(r)| = r \cdot \mathfrak{G}(i(r)) = I.$$

$$19.18: \quad [\text{ad}(a) \cdot |a| > 1 \cdot E(\mathfrak{G}(a))] \supset_a \cdot \mathfrak{G}(a) = \mathfrak{G}(a_1, \mathfrak{G}(a_2)).$$

Proofs. 19.17 is provable by induction with respect to r . $19.17 \vdash \Sigma a \cdot \text{ad}(a) \cdot |a| > 1 \cdot E(\mathfrak{G}(a))$; and, assuming $\text{ad}(a) \cdot |a| > 1 \cdot E(\mathfrak{G}(a))$, $\mathfrak{G}(a) = \mathfrak{G}(a_1, \mathfrak{G}(a_2))$ (by 19.7, 12.5, §2). Hence, by Theorem I, $\vdash 19.18$.

$$19.19: \quad \begin{aligned} N(\rho) \supset_\rho \cdot \text{ad}(a) E(\mathfrak{G}(a)) \supset_a \cdot \mathfrak{G}(a) &= \mathfrak{G}(\mathfrak{G}^\rho(a)). \\ N(\rho) \supset_\rho \cdot \text{ad}(a) E(\mathfrak{G}(\mathfrak{G}^\rho(a))) \supset_a \cdot \mathfrak{G}(a) &= \mathfrak{G}(\mathfrak{G}^\rho(a)). \end{aligned}$$

Proof. Note that $N(n)$, $19.17 \vdash \Sigma a \cdot \text{ad}(a) \cdot |a| = n \cdot E(\mathfrak{G}(a))$. Using this relation, 19.1, 19.9, 19.5, 19.7, 11.2, §2, and Theorem I, we can prove $N(r) \supset_r \cdot [\text{ad}(a) \cdot |a| = r \cdot E(\mathfrak{G}(a))] \supset_a \cdot \mathfrak{G}(a) = \mathfrak{G}(\mathfrak{G}(a))$ by induction with respect to r . Thence, using 19.17, 19.2 and Theorem I, $\text{ad}(a) E(\mathfrak{G}(a)) \supset_a \cdot \mathfrak{G}(a) = \mathfrak{G}(\mathfrak{G}(a))$. The first of the formulas 19.19 follows by induction with respect to ρ ; and the second is proved similarly.

$$19.20: \quad \text{ad}(a) \text{ad}(b) E(\mathfrak{G}(a, \mathfrak{G}(b))) \supset_{ab} \cdot \mathfrak{G}(a, \mathfrak{G}(b)) = \mathfrak{G}(\langle a, b \rangle).$$

Proof. $19.17 \vdash \Sigma ab \cdot \text{ad}(a) \text{ad}(b) E(\mathfrak{G}(a, \mathfrak{G}(b)))$. Assume $\text{ad}(a) \text{ad}(b) E(\mathfrak{G}(a, \mathfrak{G}(b)))$. Then $\mathfrak{G}(a, \mathfrak{G}(b)) = \mathfrak{G}(\mathfrak{G}^{|b|}(a), \mathfrak{G}(\mathfrak{G}^{|a|}(b)))$ (19.19, 19.2), $= \mathfrak{G}(\langle a, b \rangle)$ (19.2, 19.9, 19.5, def. of \mathfrak{G}).

$$19.21: \quad [\text{ad}(a) \text{ad}(b) E(\mathfrak{G}(a)) \cdot \Delta_b^a = 2] \supset_{ab} \cdot \mathfrak{G}(a) = \mathfrak{G}(b).$$

Proof. $19.17, 19.16 \vdash \Sigma ab \cdot \text{ad}(a) \text{ad}(b) E(\mathfrak{G}(a)) \cdot \Delta_b^a = 2$. Assuming $\text{ad}(a) \text{ad}(b) E(\mathfrak{G}(a)) \cdot \Delta_b^a = 2$, then $\mathfrak{G}(a) = \mathfrak{G}(\mathfrak{G}^{|b|}(a))$ (19.19, 19.2), $= \mathfrak{G}(\mathfrak{G}^{|a|}(b))$ (19.15, 19.14, 19.13, 19.2, 19.9), $= \mathfrak{G}(b)$.

A combination \bar{A} shall be said to be *representative* of a formula A , if $\bar{A} \text{ conv } \lambda \Pi \Sigma \& \cdot A \cdot E(\Pi)$.

19VIII. Given a formula A having no free symbols other than Π , Σ and $\&$, a representative combination \bar{A} can be found.

Proof. By C6V, there is a combination \bar{A} (in the sense of § C6) such that $\bar{A} \text{ conv } \lambda \Pi \Sigma \& \cdot A \cdot E(\Pi)$. Under the hypothesis, $\lambda \Pi \Sigma \& \cdot A \cdot E(\Pi)$ con-

tains no free symbols, and hence, by C5VI, \bar{A} is a combination in the present sense.

Let the subsequences (including the null sequence) of the sequence $\Pi, \Sigma, \&$ be X_{i1}, \dots, X_{ia_i} ($i = 1, \dots, 2^3$). By C6V and C5VI, there are combinations \mathfrak{S}_{ij} , \mathfrak{Z}_i and \mathfrak{U}_{ij} convertible into $\lambda f p \Pi \Sigma \& \cdot f(X_{i1}, \dots, X_{ia_i}, p(X_{j1}, \dots, X_{ja_j})) \cdot E(\Pi)$, $\lambda f \Pi \Sigma \& \cdot \Sigma(f(X_{i1}, \dots, X_{ia_i})) \cdot E(\Pi)$ and $\lambda f g \Pi \Sigma \& \cdot \Pi(f(X_{i1}, \dots, X_{ia_i}), g(X_{j1}, \dots, X_{ja_j})) \cdot E(\Pi)$, respectively ($i, j = 1, \dots, 2^3$).*

We denote the rules of procedure of Rosser, *loc. cit.*, Section H,† by R_1, \dots, R_{38} , and list the rules R_{ik} , "If $\mathfrak{S}_{ik}(f, p)$, then $\mathfrak{Z}_i(f)$," ($i, k = 1, \dots, 2^3$), as R_{39} - R_{102} , and the rules R_{ijk} , "If $\mathfrak{U}_{ij}(f, g)$ and $\mathfrak{S}_{ik}(f, p)$, then $\mathfrak{S}_{jk}(g, p)$," ($i, j, k = 1, \dots, 2^3$) as R_{103} - R_{614} .

19IX(t). If C is derivable from A (A and B) by an application of R_t , then $A(\Pi, \Sigma, \&)$ ($A(\Pi, \Sigma, \&), B(\Pi, \Sigma, \&)$) $\vdash C(\Pi, \Sigma, \&)$, ($t = 1, \dots, 614$).

Proof. If C is derivable from A by an application of one of R_1 - R_{38} , then A conv C . If C is derivable from A (A and B) by an application of one of the rules $R_{ik}(R_{ijk})$, then $C(\Pi, \Sigma, \&)$ is derivable from $A(\Pi, \Sigma, \&)$ ($A(\Pi, \Sigma, \&)$ and $B(\Pi, \Sigma, \&)$) by conversion, Rule IV (V) and the relations $PQ \vdash P, PQ \vdash Q$ and $P, Q \vdash PQ$.

Let $\mathfrak{A}_1, \dots, \mathfrak{A}_{13}$ be combinations representative of Axioms 1, 3-11, 14-16; respectively, and let $\alpha_1, \dots, \alpha_{13}$ be metads representing $\mathfrak{A}_1, \dots, \mathfrak{A}_{13}$, respectively (cf. 19VIII, 19IIIa).

19X. If the combination \bar{D} is representative of a formula D provable in C_1 , then \bar{D} is derivable from $\mathfrak{A}_1, \dots, \mathfrak{A}_{13}$ by means of Rules R_1 - R_{614} .

Proof. Under the hypothesis, \bar{D} is derivable from $\mathfrak{A}_1, \dots, \mathfrak{A}_{13}$ by means of conversion and the two rules

IV'. If $\lambda \Pi \Sigma \& \cdot F(P) \cdot E(\Pi)$, then $\lambda \Pi \Sigma \& \cdot \Sigma(F) \cdot E(\Pi)$.

V'. If $\lambda \Pi \Sigma \& \cdot \Pi(F, G) \cdot E(\Pi)$ and $\lambda \Pi \Sigma \& \cdot F(P) \cdot E(\Pi)$, then $\lambda \Pi \Sigma \& \cdot G(P) \cdot E(\Pi)$.

* More explicitly, let $\alpha_1 = 3, \alpha_2 = \alpha_3 = \alpha_4 = 2, \alpha_5 = \alpha_6 = \alpha_7 = 1, \alpha_8 = 0$; and let $X_{11}, X_{12}, X_{13}; X_{21}, X_{22}; X_{31}, X_{32}; X_{41}, X_{42}; X_{51}; X_{61}; X_{71}$ stand for $\Pi, \Sigma, \&; \Pi, \Sigma; \Pi, \&; \Sigma, \&; \Pi; \Sigma; \&$, respectively. Then \mathfrak{S}_{13} shall be a combination convertible into $\lambda f p \Pi \Sigma \& \cdot f(\Pi, \Sigma, \&, p(\Pi, \&)) \cdot E(\Pi)$, \mathfrak{S}_{85} a combination convertible into $\lambda f p \Pi \Sigma \& \cdot f(p(\Pi)) \cdot E(\Pi)$, etc.

† See the footnote of § C6 (*Annals of Mathematics*, vol. 35, p. 537, (12)).

If $\lambda\Pi\Sigma\&\cdot F(P)\cdot E(\Pi)$ contains no free symbols, and if $X_{i_1}, \dots, X_{i_{a_i}}$ and $X_{k_1}, \dots, X_{k_{a_k}}$ are the sets of the symbols $\Pi, \Sigma, \&$ which occur in F and P , respectively, as free symbols, then $\lambda X_{i_1} \dots X_{i_{a_i}} \cdot F$ and $\lambda X_{k_1} \dots X_{k_{a_k}} \cdot P$ contain no free symbols, and are hence convertible into combinations F' and P' , respectively (C6V, C5VI). Then $\lambda\Pi\Sigma\&\cdot F(P)\cdot E(\Pi) \text{ conv } \mathfrak{S}_{ik}(F', P')$ and $\lambda\Pi\Sigma\&\cdot \Sigma(F)\cdot E(\Pi) \text{ conv } \mathfrak{Z}_i(F')$. Hence, if A (containing no free symbols) yields C by an application of IV', then C is derivable from A by conversion and an application of one of the rules R_{ik} in which the premise and conclusion are combinations. Similarly, if A and B (containing no free symbols) yield C by an application of V', then C is derivable from A and B by conversion and an application of one of the rules R_{ijk} in which the premise and conclusion are combinations. The formulas derivable from $\mathfrak{A}_1, \dots, \mathfrak{A}_{13}$ by conversion, IV' and V' contain no free symbols (cf. C5V Cor.). Hence \bar{D} is derivable from $\mathfrak{A}_1, \dots, \mathfrak{A}_{13}$ by conversion and applications of R_{ik} and R_{ijk} in which the premises and conclusions are combinations. Now R_1 - R_{38} have the property that if A and C are combinations, and $A \text{ conv } C$, then C is derivable from A by R_1 - R_{38} . Hence \bar{D} is derivable from $\mathfrak{A}_1, \dots, \mathfrak{A}_{13}$ by R_1 - R_{38} , R_{ik} , R_{ijk} , *i. e.* by R_1 - R_{614} .

We now define expressions \mathfrak{R}_t corresponding to the rules R_t ($t = 1, \dots, 614$).

For typical rules of the set R_1 - R_{38} , the definition of \mathfrak{R}_t follows (r_t standing for an expression satisfying the condition $r_t(1) \text{ conv } I$ and the condition given below):

R_1 . If $I(p)$, then p .

$$r_1(2) \text{ conv } \lambda a \cdot a_2, \quad \mathfrak{R}_1 \rightarrow \lambda a \cdot r_1(\epsilon_1^{|a|} \circ \Delta_{[1]}^{a_1}, a).$$

R_2 . If p , then $I(p)$.

$$\mathfrak{R}_2 \rightarrow \lambda a \cdot \langle [1], a \rangle.$$

R_3 . If $f(I(p, q))$, then $f(p(q))$.

$$r_3(2) \text{ conv } \lambda a \cdot \langle a_1, \langle a_{212}, a_{22} \rangle \rangle, \quad \mathfrak{R}_3 \rightarrow \lambda a \cdot r_3(\epsilon_3^{|a|} \circ \Delta_{[1]}^{a_{211}}, a).$$

R_6 . If $f(p(q, p(s, r)))$, then $f(J(p, q, r, s))$.

$$r_6(2) \text{ conv } \lambda a \cdot \langle a_1, \langle \langle \langle [2], a_{211} \rangle, a_{212} \rangle, a_{222} \rangle, a_{2212} \rangle \rangle, \\ \mathfrak{R}_6 \rightarrow \lambda a \cdot r_6(\epsilon_6^{|a|} \circ \Delta_{a_{2211}}^{a_{211}}, a).^*$$

If R_t is the rule R_{ik} (for a certain i and k), then \mathfrak{R}_t shall be the expres-

* The considerations governing the choice of the \mathfrak{R}_t will appear in the proofs of *19XI(t) and 19.23(t). a_2, a_{211}, \dots are our abbreviations for. $\mathfrak{M}_2(a), \mathfrak{M}_1(\mathfrak{M}_1(\mathfrak{M}_2(a))), \dots$

sion \mathfrak{R}_{ik} defined thus: Let \mathfrak{s}_{ik} and \mathfrak{t}_i be metads which represent the combinations \mathfrak{S}_{ik} and \mathfrak{T}_i , respectively (cf. 19IIIa), and let \mathfrak{r}_{ik} be an expression such that $\mathfrak{r}_{ik}(1) \text{ conv } I$ and $\mathfrak{r}_{ik}(2) \text{ conv } \lambda a \cdot \langle \mathfrak{t}_i, a_{12} \rangle$. Let $\mathfrak{R}_{ik} \rightarrow \lambda a \cdot \mathfrak{r}_{ik}(\epsilon_2^{[a]} \circ \Delta(a_{11}, \mathfrak{s}_{ik}), a)$.

If R_t is the rule R_{ijk} (for a certain i, j and k), then \mathfrak{R}_t shall be the expression \mathfrak{R}_{ijk} defined thus: Let \mathfrak{u}_{ij} be a metad which represents \mathfrak{U}_{ij} , and \mathfrak{r}_{ijk} an expression such that $\mathfrak{r}_{ijk}(1) \text{ conv } \lambda pq \cdot I^{[p]}(q)$ and $\mathfrak{r}_{ijk}(2) \text{ conv } \lambda ab \cdot \langle \langle \mathfrak{s}_{jk}, a_2 \rangle, b_2 \rangle$. Let $\mathfrak{R}_{ijk} \rightarrow \lambda ab \cdot \mathfrak{r}_{ijk}(\epsilon_2^{[a]} \circ \epsilon_2^{[b]} \circ \Delta(a_{11}, \mathfrak{u}_{ij}) \circ \Delta(b_{11}, \mathfrak{s}_{jk}) \circ \Delta_{b_{12}}^{a_{12}}, a, b)$.

19XI(t). If the metad \mathfrak{a} represents (the metads $\mathfrak{a}, \mathfrak{b}$ represent) a combination \mathfrak{A} (combinations $\mathfrak{A}, \mathfrak{B}$) such that R_t is applicable to \mathfrak{A} (to the pair $\mathfrak{A}, \mathfrak{B}$), then $\mathfrak{R}_t(\mathfrak{a})(\mathfrak{R}_t(\mathfrak{a}, \mathfrak{b}))$ is a metad which represents the combination resulting from the application. ($t = 1, \dots, 614$).

As illustrative of the arguments for the several values of t , we take the case of a $t > 102$ (≤ 614). Then R_t is the rule R_{ijk} , for a certain i, j and k ; and \mathfrak{A} and \mathfrak{B} are of the forms $\mathfrak{U}_{ij}(\mathfrak{f}, \mathfrak{g})$ and $\mathfrak{S}_{ik}(\mathfrak{f}, \mathfrak{p})$, respectively ($\mathfrak{f}, \mathfrak{g}$ and \mathfrak{p} being combinations, by C6II). Then the ranks of \mathfrak{A} and \mathfrak{B} are both at least 3. Hence, by 15Ik, 19IIIb and 19I, $\epsilon_2^{[a]} \text{ conv } \epsilon_2^{[b]} \text{ conv } 2$. \mathfrak{u}_{ij} , \mathfrak{s}_{ik} and \mathfrak{s}_{jk} are, by definition, metads which represent the combinations \mathfrak{U}_{ij} , \mathfrak{S}_{ik} , and \mathfrak{S}_{jk} , respectively. Also, by 19IIIc, the metads \mathfrak{a}_{11} , \mathfrak{a}_{12} , \mathfrak{a}_2 , \mathfrak{b}_{11} , \mathfrak{b}_{12} , \mathfrak{b}_2 represent the combinations \mathfrak{U}_{ij} , $\mathfrak{f}, \mathfrak{g}, \mathfrak{S}_{ik}, \mathfrak{f}, \mathfrak{p}$, respectively. Hence, by 19VI, $\Delta(\mathfrak{a}_{11}, \mathfrak{u}_{ij}) \text{ conv } \Delta(\mathfrak{b}_{11}, \mathfrak{s}_{ik}) \text{ conv } \Delta_{b_{12}}^{a_{12}} \text{ conv } 2$. Then, by 15Ij, $\epsilon_2^{[a]} \circ \epsilon_2^{[b]} \circ \Delta(\mathfrak{a}_{11}, \mathfrak{u}_{ij}) \circ \Delta(\mathfrak{b}_{11}, \mathfrak{s}_{ik}) \circ \Delta_{b_{12}}^{a_{12}} \text{ conv } 2$. Consequently $\mathfrak{R}_{ijk}(\mathfrak{a}, \mathfrak{b}) \text{ conv } \mathfrak{r}_{ijk}(2, \mathfrak{a}, \mathfrak{b})$, $\text{conv } \langle \langle \mathfrak{s}_{jk}, \mathfrak{a}_2 \rangle, \mathfrak{b}_2 \rangle$. By 19V, the latter is a metad which represents $\mathfrak{S}_{jk}(\mathfrak{g}, \mathfrak{p})$, which is the formula resulting from the application of R_{ijk} to $\mathfrak{A}, \mathfrak{B}$.

COROLLARY. If the combination $\bar{\mathfrak{D}}$ is derivable from $\mathfrak{U}_1, \dots, \mathfrak{U}_{13}$ by R_1 - R_{614} , the set of formulas derivable from $\mathfrak{a}_1, \dots, \mathfrak{a}_{13}$ by zero or more operations of passing from $\mathfrak{a}, \mathfrak{b}$ to $\mathfrak{R}_1(\mathfrak{a}), \dots, \mathfrak{R}_{102}(\mathfrak{a}), \mathfrak{R}_{103}(\mathfrak{a}, \mathfrak{b}), \dots$, or $\mathfrak{R}_{614}(\mathfrak{a}, \mathfrak{b})$ contains a metad which represents $\bar{\mathfrak{D}}$.

This follows from the Theorems 19XI(t) by the definition of $\mathfrak{a}_1, \dots, \mathfrak{a}_{13}$ as metads representing the combinations $\mathfrak{U}_1, \dots, \mathfrak{U}_{13}$, respectively.

Now let \mathfrak{S} be an expression which has the properties (1) and (2) of \mathfrak{H} in 17II when $\mathfrak{A}_1, \dots, \mathfrak{A}_l, \mathfrak{R}_1, \dots, \mathfrak{R}_{m+n}$, m, n are taken to be $\mathfrak{a}_1, \dots, \mathfrak{a}_{13}, \mathfrak{R}_1, \dots, \mathfrak{R}_{614}, 102, 512$, respectively.

19XII. If the combination $\bar{\mathfrak{D}}$ is representative of a formula \mathfrak{D} provable

in C_1 , then there is a positive integer n such that $\mathfrak{S}(n)$ is a metad which represents \bar{D} .

Proof. By 19X, \bar{D} is derivable from $\mathfrak{A}_1, \dots, \mathfrak{A}_{13}$ by R_1 - R_{614} . The conclusion follows by 19XI Cor. and 17II(1) (under our definition of \mathfrak{S}).

Let $G \rightarrow \lambda a \cdot \mathfrak{G}(a, \Pi, \Sigma, \&)$.

$$19.22(s) \quad \text{ad}(\mathfrak{a}_s)G(\mathfrak{a}_s) \quad (s = 1, \dots, 13).$$

Proof. Since \mathfrak{a}_s is a given metad, $\text{ad}(\mathfrak{a}_s)$ is provable from the formulas 19.1 by a succession of applications of 19.3. Since \mathfrak{a}_s represents the combination \mathfrak{A}_s , which is representative of an axiom A_s , $G(\mathfrak{a}_s) \text{ conv } \mathfrak{G}(\mathfrak{a}_s, \Pi, \Sigma, \&)$, $\text{conv } \mathfrak{A}_s(\Pi, \Sigma, \&)$ (19VII), $\text{conv } \{\lambda \Pi \Sigma \& \cdot A_s \cdot E(\Pi)\}(\Pi, \Sigma, \&)$, $\text{conv } A_s \cdot E(\Pi)$, which is a provable formula.

$$19.23(t): \quad \begin{aligned} & \text{ad}(a)G(a) \supset_a \text{ad}(\mathfrak{N}_t(a))G(\mathfrak{N}_t(a)) \quad (t = 1, \dots, 102). \\ & [\text{ad}(a)G(a) \cdot \text{ad}(b)G(b)] \supset_{ab} \text{ad}(\mathfrak{N}_t(a, b))G(\mathfrak{N}_t(a, b)) \\ & \quad (t = 103, \dots, 614). \end{aligned}$$

Proof. We take as typical the case of a $t > 102$. Then \mathfrak{N}_t is one of the expressions \mathfrak{N}_{ijk} for a certain i, j and k . 19.22 $\vdash \Sigma ab \cdot \text{ad}(a)G(a) \cdot \text{ad}(b)G(b)$. Assume $\text{ad}(a)G(a) \cdot \text{ad}(b)G(b)$. Since \mathfrak{u}_{ij} , \mathfrak{s}_{ik} and \mathfrak{s}_{jk} are given metads, $\text{ad}(\mathfrak{u}_{ij})$, $\text{ad}(\mathfrak{s}_{ik})$ and $\text{ad}(\mathfrak{s}_{jk})$ are provable. Using 14.2, 14.7, 19.2, 19.8, and 19.12, $M(\epsilon_2^{|a|} \circ \epsilon_2^{|b|} \circ \Delta(a_{11}, \mathfrak{u}_{ij}) \circ \Delta(b_{11}, \mathfrak{s}_{ik}) \circ \Delta_{b_{12}}^{a_{12}})$. Case 1: $\epsilon_2^{|a|} \circ \epsilon_2^{|b|} \circ \Delta(a_{11}, \mathfrak{u}_{ij}) \circ \Delta(b_{11}, \mathfrak{s}_{ik}) \circ \Delta_{b_{12}}^{a_{12}} = 1$. Then $\mathfrak{N}_{ijk}(a, b) = \mathfrak{r}_{ijk}(1, a, b)$, $\text{conv } I^{|a|}(b) = b$ (19.2, 7.2), and $\text{ad}(\mathfrak{N}_{ijk}(a, b))G(\mathfrak{N}_{ijk}(a, b))$ follows from $\text{ad}(b)G(b)$. Case 2: $\epsilon_2^{|a|} \circ \epsilon_2^{|b|} \circ \Delta(a_{11}, \mathfrak{u}_{ij}) \circ \Delta(b_{11}, \mathfrak{s}_{ik}) \circ \Delta_{b_{12}}^{a_{12}} = 2$. Then $\mathfrak{N}_{ijk}(a, b) = \mathfrak{r}_{ijk}(2, a, b)$, $\text{conv } \langle \langle \mathfrak{s}_{jk}, a_2 \rangle, b_2 \rangle$, and $\text{ad}(\mathfrak{N}_{ijk}(a, b))$ follows from $\text{ad}(\mathfrak{s}_{jk})$, $\text{ad}(a)$ and $\text{ad}(b)$ by means of 19.8 and 19.11. Also $|a| > 2$, $|b| > 2$, $\Delta(a_{11}, \mathfrak{u}_{ij}) = 2$, $\Delta(b_{11}, \mathfrak{s}_{ik}) = 2$, $\Delta_{b_{12}}^{a_{12}} = 2$ (14.2, 14.6, 14.7, 14.9, 19.2, 19.8, 19.12). Now, from $G(a)$ by conversion, $\mathfrak{G}(a, \Pi, \Sigma, \&)$; thence, by 19.18, $\mathfrak{G}(a_1, \mathfrak{G}(a_2), \Pi, \Sigma, \&)$; by another application of 19.18, $\mathfrak{G}(a_{11}, \mathfrak{G}(a_{12}), \mathfrak{G}(a_2), \Pi, \Sigma, \&)$; by two applications of 19.21, $\mathfrak{G}(\mathfrak{u}_{ij}, \mathfrak{G}(b_{12}), \mathfrak{G}(a_2), \Pi, \Sigma, \&)$; and, by conversion (cf. 19VII), $\mathfrak{A}(\Pi, \Sigma, \&)$, where $\mathfrak{A} \rightarrow \mathfrak{u}_{ij}(\mathfrak{G}(b_{12}), \mathfrak{G}(a_2))$. Similarly, from $G(b)$ we infer $\mathfrak{B}(\Pi, \Sigma, \&)$, where $\mathfrak{B} \rightarrow \mathfrak{s}_{ik}(\mathfrak{G}(b_{12}), \mathfrak{G}(b_2))$. If $\mathfrak{C} \rightarrow \mathfrak{s}_{jk}(\mathfrak{G}(a_2), \mathfrak{G}(b_2))$, then \mathfrak{C} is derivable from \mathfrak{A} and \mathfrak{B} by an application of R_{ijk} . Hence, by 19IX, we can infer $\mathfrak{C}(\Pi, \Sigma, \&)$ from $\mathfrak{A}(\Pi, \Sigma, \&)$ and $\mathfrak{B}(\Pi, \Sigma, \&)$. From $\mathfrak{C}(\Pi, \Sigma, \&)$, by conversion, $\mathfrak{G}(\mathfrak{s}_{jk}, \mathfrak{G}(a_2), \mathfrak{G}(b_2), \Pi, \Sigma, \&)$ (19VII); by applications of 19.20,

$\mathfrak{G}(\langle \langle \mathfrak{s}_{jk}, a_2 \rangle, b_2 \rangle, \Pi, \Sigma, \&)$; by conversion, $G(\langle \langle \mathfrak{s}_{jk}, a_2 \rangle, b_2 \rangle)$; and thence $G(\mathfrak{R}_{ijk}(a, b))$. Using Axiom 14, $\text{ad}(\mathfrak{R}_{ijk}(a, b)) \cdot G(\mathfrak{R}_{ijk}(a, b))$. By cases (C9I), $\text{ad}(\mathfrak{R}_{ijk}(a, b)) G(\mathfrak{R}_{ijk}(a, b))$.

19.24: $N(n) \supset_n \text{ad}(\mathfrak{S}(n)) G(\mathfrak{S}(n))$.

This formula follows from the formulas 19.22(s) and 19.23(t) by 17II(2) and our definition of \mathfrak{S} .

19XIII. If $\mathbf{F}(\mathbf{P})$ is provable in C_1 , and \mathbf{P} contains no free symbols, then a formula \mathbf{U} (containing no free symbols) can be found such that (1) if $\mathbf{F}(\mathbf{Q})$ is provable in C_1 , and \mathbf{Q} contains no free symbols, then there is a positive integer q such that $\mathbf{U}(q) \text{ conv } \mathbf{Q}$, and (2) $N(n) \supset_n \mathbf{F}(\mathbf{U}(n))$ is provable.

Proof. Assume the hypothesis. Let \mathbf{F}' and \mathbf{P}' be combinations such that $\mathbf{F}' \text{ conv } \lambda p \Pi \Sigma \& \cdot \mathbf{F}(p) \cdot E(\Pi)$, and $\mathbf{P}' \text{ conv } \mathbf{P}$ (C6V, C5VI, C5V Cor.). Let \mathbf{c} be a metad representing $\mathbf{F}'(\mathbf{P}')$ (19IIIa). Let \mathbf{K} be an expression such that $\mathbf{K}(1) \text{ conv } \lambda a \cdot I^{|\mathbf{a}|}(\mathbf{c})$ and $\mathbf{K}(2) \text{ conv } I$. Let $\mathbf{L} \rightarrow \lambda a \cdot \mathbf{K}(\epsilon_1^{|\mathbf{a}|} \circ \Delta_{c_1}^{a_1}, a)$.

(1) If the metad \mathbf{a} represents a combination of the form $\mathbf{F}'(\mathbf{Q}')$, then $\mathbf{L}(\mathbf{a}) \text{ conv } \mathbf{a}$ (15Ij, k, 19I, 19IIIb, f, 19VI).

(2) $\vdash \text{ad}(\mathbf{a}) G(\mathbf{a}) \supset_a \text{ad}(\mathbf{L}(\mathbf{a})) G(\mathbf{L}(\mathbf{a})) \cdot \epsilon(|\mathbf{L}(\mathbf{a})|, 1) \circ \Delta(\mathbf{L}(\mathbf{a})_1, \mathbf{c}_1) = 2$.

Proof. Assume $\text{ad}(\mathbf{a}) G(\mathbf{a})$. Case 1: $\epsilon_1^{|\mathbf{a}|} \circ \Delta_{c_1}^{a_1} = 1$. Then $\mathbf{L}(\mathbf{a}) = \mathbf{K}(1, \mathbf{a})$, $= \mathbf{c}$ (19.2, 7.2). 19.1, 19.3 $\vdash \text{ad}(\mathbf{c})$; $G(\mathbf{c})$ is provable by conversion from $\mathbf{F}(\mathbf{P}) \cdot E(\Pi)$; and $\epsilon_1^{|\mathbf{c}|} \circ \Delta_{c_1}^{c_1} \text{ conv } 2$. Case 2: $\epsilon_1^{|\mathbf{a}|} \circ \Delta_{c_1}^{a_1} = 2$. Then $\mathbf{L}(\mathbf{a}) = \mathbf{K}(2, \mathbf{a})$, $\text{conv } \mathbf{a}$. In both cases $\text{ad}(\mathbf{L}(\mathbf{a})) G(\mathbf{L}(\mathbf{a})) \cdot \epsilon(|\mathbf{L}(\mathbf{a})|, 1) \circ \Delta(\mathbf{L}(\mathbf{a})_1, \mathbf{c}_1) = 2$ is provable from the assumptions; and hence, by applications of C9I and Theorem I, (2) holds.

Let $\mathfrak{B} \rightarrow \lambda a \cdot \mathfrak{G}(a_2)$.

(3) If the metad \mathbf{a} represents a combination of the form $\mathbf{F}'(\mathbf{Q}')$, then $\mathfrak{B}(\mathbf{a}) \text{ conv } \mathbf{Q}'$ (19III f, 19VII).

(4) $\vdash [\text{ad}(\mathbf{a}) G(\mathbf{a}) \cdot \epsilon_1^{|\mathbf{a}|} \circ \Delta_{c_1}^{a_1} = 2] \supset_a F(\mathfrak{B}(\mathbf{a}))$.

Proof. By 19.22 and (2), $\Sigma a \cdot \text{ad}(\mathbf{a}) G(\mathbf{a}) \cdot \epsilon_1^{|\mathbf{a}|} \circ \Delta_{c_1}^{a_1} = 2$. Assume $\text{ad}(\mathbf{a}) G(\mathbf{a}) \cdot \epsilon_1^{|\mathbf{a}|} \circ \Delta_{c_1}^{a_1} = 2$. Then $|\mathbf{a}| > 1$ and $\Delta_{c_1}^{a_1} = 2$ (14.6, 14.7, 14.9, 19.2, 19.8, 19.12). Now $G(\mathbf{a}) \text{ conv } \mathfrak{G}(a, \Pi, \Sigma, \&)$; thence, by 19.18, $\mathfrak{G}(a_1, \mathfrak{G}(a_2), \Pi, \Sigma, \&)$; by 19.21, $\mathfrak{G}(\mathbf{c}_1, \mathfrak{G}(a_2), \Pi, \Sigma, \&)$. $\mathfrak{G}(\mathbf{c}_1, \mathfrak{G}(a_2), \Pi, \Sigma, \&)$

conv $F'(\mathfrak{G}(a_2), \Pi, \Sigma, \&)$ (19III \bar{f} , 19VII), conv $F(\mathfrak{G}(a_2)) \cdot E(\Pi)$ (def. of F'), conv $F(\mathfrak{B}(a)) \cdot E(\Pi)$, whence, by Axiom 15, $F(\mathfrak{B}(a))$.

Let $U \rightarrow \lambda n \cdot \mathfrak{B}(L(\mathfrak{H}(n)))$.

(5) Suppose that $F(Q)$ is provable in C_1 , and that Q contains no free symbols. Let Q' be a combination such that $Q' \text{ conv } Q$. Then the combination $F'(Q')$ is representative of $F(Q)$. Hence, by 19XII, there is a positive integer q such that $\mathfrak{H}(q)$ represents $F'(Q')$. Now $U(q) \text{ conv } \mathfrak{B}(L(\mathfrak{H}(q)))$, conv $\mathfrak{B}(\mathfrak{H}(q))$ (by (1)), conv Q' (by (3)), conv Q .

(6) Assume $N(n)$. By 19.24, $\text{ad}(\mathfrak{H}(n)) \cdot G(\mathfrak{H}(n))$. Thence, using (2) and (4), $F(\mathfrak{B}(L(\mathfrak{H}(n))))$, and, by conversion, $F(U(n))$. By Theorem I, $N(n) \supset_n F(U(n))$.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

DOUBLY PERIODIC FUNCTIONS OF THE SECOND KIND AND THE ARITHMETICAL FORM $xy + zn$.

By E. T. BELL.

1. *Introduction.* The sixteen doubly periodic functions of the second kind,

$$\phi_{abc}(x, y) \equiv \vartheta'_1 \vartheta_a(x + y) / \vartheta_b(x) \vartheta_c(y),$$

where the triple index abc has the values

$$\begin{array}{cccc} 001, & 010, & 023, & 032, \\ 100, & 111, & 122, & 133, \\ 203, & 212, & 221, & 230, \\ 302, & 313, & 320, & 331, \end{array}$$

give rise to a set of identities of the form

$$(1) \quad \phi_{abc}(x, y) \phi_{rst}(x, -y) \equiv AB + CD,$$

where each of A, B, C, D is a function of x alone, or of y alone, on reducing the numerator $\vartheta'_1 \vartheta_a(x + y) \vartheta_r(x - y)$ of the left by means of the addition formulas for the thetas. An identity (1) is said to be of the second degree (with reference to the right) in theta quotients provided that each of A, B, C, D has a Fourier expansion in which the coefficients of the several powers of q involve only functions of the divisors of the exponents.

The complete set of identities (1) of the second degree contains a subset of identities from which the entire set can be generated by transformations of the forms

$$(2) \quad q \rightarrow -q, \quad x \rightarrow x \pm \pi/2, \quad y \rightarrow y \pm \pi/2; \\ (x, y) \rightarrow (y, x), \quad (x, y) \rightarrow (x, -y), \quad (x, y) \rightarrow (-x, y),$$

or by repetitions of these, and no identity in the subset is obtainable by these transformations from any other in the subset. It is easily seen that this subset contains precisely 25 identities. By the method of paraphrase, each of these identities implies and is implied by an arithmetical identity concerning parity functions summed over a quadratic partition. The set of 25 arithmetical identities (given in § 4) is thus equivalent to all the identities of the type (1), of the second degree, obtainable from the doubly periodic functions of the second kind, since transformations of the type (2) do not increase or diminish the generality of a parity identity (if the transformations are applied to the trigonometric identity, to which a particular identity (1) is equivalent, before

paraphrasing). In § 5, arithmetical identities of a new, completely general type are indicated.

Let the functions $f(x, y)$, $g(x, y)$ be single-valued and finite for all pairs of integer values of x, y , and beyond the parity conditions

$$(3) \quad f(x, y) = f(-x, -y); \quad g(x, y) = -g(-x, -y), \quad g(0, 0) = 0,$$

let f, g be entirely arbitrary. In the notation of parity functions,

$$(4) \quad f(x, y) \equiv f((x, y) |), \quad g(x, y) \equiv g(| (x, y));$$

the parity of f is $(2 | 0)$, that of g is $(0 | 2)$. In an identity involving functions f or g with integer arguments x, y , all the (x, y) have the same character $(x_0, y_0) \bmod 2$, namely, $x \equiv x_0, y \equiv y_0 \bmod 2$. Hence in any such identity we may replace $f(x, y)$, $g(x, y)$ by the functions indicated next, since the transformed functions have the same respective parities as f, g :

$$\begin{aligned} x \equiv 0 \bmod 2: f(x, y) &\rightarrow (-1)^{x/2} f(x, y), & g(x, y) &\rightarrow (-1)^{x/2} g(x, y); \\ y \equiv 0 \bmod 2: f(x, y) &\rightarrow (-1)^{y/2} f(x, y), & g(x, y) &\rightarrow (-1)^{y/2} g(x, y); \\ x \equiv 1 \bmod 2: f(x, y) &\rightarrow (-1 | x) g(x, y), & g(x, y) &\rightarrow (-1 | x) f(x, y); \\ y \equiv 1 \bmod 2: f(x, y) &\rightarrow (-1 | y) g(x, y), & g(x, y) &\rightarrow (-1 | y) f(x, y); \\ f(x, y) &\rightarrow f(ax, by), & g(x, y) &\rightarrow g(ax, by), \end{aligned}$$

where $(-1 | x)$ is defined only for odd integers x , and is $(-1)^{(x-1)/2}$, and a, b are arbitrary integers different from zero. These transformations, or others compounded from them, are called the elementary transformations of f, g .

The set of 25 arithmetical identities described above is such that no identity in the set is obtainable from any other by elementary transformations of the functions f, g . The partitions concerned are all of the form $xy + zw$, where x, y, z, w are non-negative integers. With respect to elementary transformations this set of 25 is the irreducible equivalent of the entire set of identities (1) of the second degree. They are obtained from the subset defined in connection with the transformations (2).

2. *Theta identities.* To write out the 25 identities mentioned at the end of § 1 we shall need the following theta quotients.

$$\begin{aligned} \psi_{10}(x) &\equiv \partial_2 \partial_3 \partial_1(x) / \partial_0(x), & \psi_{02}(x) &\equiv \partial_0 \partial_2 \partial_0(x) / \partial_2(x), \\ \psi_{20}(x) &\equiv \partial_0 \partial_2 \partial_2(x) / \partial_0(x), & \psi_{12}(x) &\equiv \partial_0 \partial_2 \partial_1(x) / \partial_2(x), \\ \psi_{30}(x) &\equiv \partial_0 \partial_3 \partial_3(x) / \partial_0(x), & \psi_{32}(x) &\equiv \partial_2 \partial_3 \partial_3(x) / \partial_2(x); \\ \psi_{01}(x) &\equiv \partial_2 \partial_3 \partial_0(x) / \partial_1(x), & \psi_{03}(x) &\equiv \partial_0 \partial_3 \partial_0(x) / \partial_3(x), \\ \psi_{21}(x) &\equiv \partial_0 \partial_3 \partial_2(x) / \partial_1(x), & \psi_{13}(x) &\equiv \partial_0 \partial_2 \partial_1(x) / \partial_3(x), \\ \psi_{31}(x) &\equiv \partial_0 \partial_2 \partial_3(x) / \partial_1(x), & \psi_{23}(x) &\equiv \partial_2 \partial_3 \partial_2(x) / \partial_3(x); \\ \chi_{0123}(x) &\equiv \partial_0^2 \partial_0(x) \partial_1(x) / \partial_2(x) \partial_3(x), \\ \chi_{0213}(x) &\equiv \partial_3^2 \partial_0(x) \partial_2(x) / \partial_1(x) \partial_3(x), \end{aligned}$$

$$\chi_{0312}(x) \equiv \vartheta_2^2 \vartheta_0(x) \vartheta_3(x) / \vartheta_1(x) \vartheta_2(x),$$

$$\chi_{1203}(x) \equiv \vartheta_2^2 \vartheta_1(x) \vartheta_2(x) / \vartheta_0(x) \vartheta_3(x),$$

$$\chi_{1302}(x) \equiv \vartheta_3^2 \vartheta_1(x) \vartheta_3(x) / \vartheta_0(x) \vartheta_2(x),$$

$$\chi_{2301}(x) \equiv \vartheta_0^2 \vartheta_2(x) \vartheta_3(x) / \vartheta_0(x) \vartheta_1(x).$$

The 25 identities of the second degree are as follows:

- (I) $\phi_{100}(x, y) \phi_{203}(x, -y) = \psi_{10}(x) \psi_{20}(x) + \psi_{30}(x) \chi_{1203}(y).$
- (II) $-\phi_{100}(x, y) \phi_{001}(x, -y) = \psi_{20}(x) \psi_{30}(x) + \psi_{10}(x) \chi_{2301}(y).$
- (III) $\phi_{100}(x, y) \phi_{302}(x, -y) = \psi_{10}(x) \psi_{30}(x) + \psi_{20}(x) \psi_{1302}(y).$
- (IV) $-\phi_{100}(x, y) \phi_{331}(x, -y) = \psi_{10}(x) \psi_{21}(y) + \psi_{23}(x) \psi_{30}(y).$
- (V) $\phi_{100}(x, y) \phi_{032}(x, -y) = \psi_{13}(x) \psi_{30}(y) + \psi_{20}(x) \psi_{12}(y).$
- (VI) $\phi_{100}(x, y) \phi_{212}(x, -y) = \psi_{20}(x) \psi_{32}(y) + \psi_{31}(x) \psi_{10}(y).$
- (VII) $\phi_{001}(x, y) \phi_{302}(x, -y) = \psi_{10}(x) \psi_{20}(x) + \psi_{30}(x) \chi_{0312}(y).$
- (VIII) $-\phi_{001}(x, y) \phi_{331}(x, -y) = \psi_{01}(y) \psi_{31}(y) + \psi_{21}(y) \chi_{1203}(x).$
- (IX) $\phi_{001}(x, y) \phi_{320}(x, -y) = \psi_{10}(x) \psi_{20}(y) + \psi_{32}(x) \psi_{31}(y).$
- (X) $-\phi_{001}(x, y) \phi_{313}(x, -y) = \psi_{20}(x) \psi_{23}(y) + \psi_{31}(x) \psi_{01}(y).$
- (XI) $-\phi_{001}(x, y) \phi_{111}(x, -y) = \psi_{21}(y) \psi_{31}(y) - \psi_{01}(y) \chi_{2301}(x).$
- (XII) $\phi_{001}(x, y) \phi_{212}(x, -y) = \psi_{21}(x) \psi_{01}(y) + \psi_{30}(x) \psi_{32}(y).$
- (XIII) $\phi_{001}(x, y) \phi_{122}(x, -y) = \psi_{12}(x) \psi_{31}(y) - \psi_{30}(x) \psi_{02}(y).$
- (XIV) $-\phi_{001}(x, y) \phi_{221}(x, -y) = \psi_{01}(y) \psi_{21}(y) + \psi_{31}(y) \chi_{1302}(x).$
- (XV) $\phi_{111}(x, y) \phi_{212}(x, -y) = \psi_{01}(x) \psi_{31}(x) + \psi_{21}(x) \chi_{0312}(y).$
- (XVI) $\phi_{100}(x, y) \phi_{100}(x, -y) = \psi_{10}^2(x) - \psi_{10}^2(y).$
- (XVII) $\phi_{100}(x, y) \phi_{133}(x, -y) = \psi_{03}(x) \psi_{30}(y) - \psi_{30}(x) \psi_{03}(y).$
- (XVIII) $-\phi_{100}(x, y) \phi_{111}(x, -y) = \psi_{10}(x) \psi_{01}(y) - \psi_{01}(x) \psi_{10}(y).$
- (XIX) $\phi_{100}(x, y) \phi_{122}(x, -y) = \psi_{02}(x) \psi_{20}(y) - \psi_{20}(x) \psi_{02}(y).$
- (XX) $-\phi_{001}(x, y) \phi_{001}(x, -y) = \psi_{20}^2(x) + \psi_{31}^2(y).$
- (XXI) $\phi_{001}(x, y) \phi_{032}(x, -y) = \psi_{03}(x) \psi_{21}(y) + \psi_{31}(x) \psi_{12}(y).$
- (XXII) $\phi_{001}(x, y) \phi_{010}(x, -y) = \psi_{01}(x) \psi_{01}(y) - \psi_{10}(x) \psi_{10}(y).$
- (XXIII) $\phi_{001}(x, y) \phi_{023}(x, -y) = \psi_{02}(x) \psi_{31}(y) + \psi_{20}(x) \psi_{13}(y).$
- (XXIV) $-\phi_{111}(x, y) \phi_{111}(x, -y) = \psi_{01}^2(y) - \psi_{01}^2(x).$
- (XXV) $\phi_{111}(x, y) \phi_{122}(x, -y) = \psi_{12}(x) \psi_{21}(y) - \psi_{21}(x) \psi_{12}(y).$

3. *Notation.* The letters m, n, d, δ, t, τ , with or without suffixes or accents, denote integers greater than zero; the n, d, δ, t may be odd or even; the m, τ are always odd. Letters m, n without suffixes denote constants; with suffixes, variables. In referring to previous papers in which parts of this notation were used, it is to be noted that if $n = 2^a m$, $a \geq 0$, the separation $n = 2^a d \delta$ is identical with the separation $n = t \tau$; namely, either $(2^a d, \delta) = (t, \tau)$ or $(2^a \delta, d) = (t, \tau)$. Similarly for accented letters, or letters with suffixes.

To paraphrase the 25 identities into their arithmetical equivalents we shall need the reduced forms of the Fourier expansions of the theta quotients

on the right of (I)-(XXV). These are given in a previous paper,* together with many more useful in similar work. The series for the ψ are in § 14, p. 172; those for the χ in § 15, p. 173; and those for the ψ^2 in § 16, p. 173, of the paper cited. The only correct list in print for the ϕ is that in § 11 of another paper.† The trigonometric identities in § 8 of that paper are used in reducing products involving sec, csc, tan, ctn to sums of sines or cosines (plus possibly a term in sec, etc.). These expansions and formulas being readily available in the papers cited, we shall not reproduce them here. It will suffice to state only the final results (all of which have been checked), as the method of paraphrase is straightforward and entirely elementary (see the second paper cited). The arithmetical equivalent of a particular identity in § 2 is numbered correspondingly; f, g are as in § 1 (3), and summations refer to all values of the variables (also to the specified divisors of the constants) in the partitions indicated in each instance.

One detail in reading the identities may be noted. In (II), for example, the outer Σ (without limits) on the right refers to all t, τ defined by the given partition, and so in all similar cases. By introducing appropriate functions of divisors, as $\zeta'_0(n)$ in (VIII), for example, reductions of such sums are sometimes possible. However, it is usually simpler to leave the identities without such reductions.

4. *Parity identities.* For the m, n, d, δ, t, τ notation see § 3. The (d, δ) and the (t, τ) , with or without suffixes or accents, denote pairs of conjugate divisors, and a particular pair refers to the m or n in the stated partition that has the same display of suffixes and accents as those in the particular pair. For example, if the partition is $n = m_i + n_j$, and the pairs $(t_i, \tau_i), (d_j, \delta_j), (d, \delta)$ occur in the parity identity, (t_i, τ_i) refer to m_i , (d_j, δ_j) to n_j , (d, δ) to n . Thus, written in full, the partition would be $n = m_i + n_j, n = d\delta, m_i = t_i\tau_i, n_j = d_j\delta_j$. This convention saves much space. Notice in particular that if the partition contains numerical factors, as in $an = bm_i + cn_j$, where a, b, c are definite integers, the $(d, \delta), (t_i, \tau_i), (d_j, \delta_j)$ refer to the divisors of n, m_i, n_j , and *not* to those of an, bm_i, cn_j . Note that the pairs of conjugates are (d, δ) and (t, τ) ; $(d, t), (d, \tau), (\delta, t), (\delta, \tau)$ do not occur.

$$(I_1) \quad 2m = m_1 + m_2; \quad m = 2n_3 + m_4:$$

$$\Sigma(-1 | \tau_2) [g(t_1 + t_2, \tau_1 - \tau_2) + g(t_1 - t_2, \tau_1 + \tau_2) - g(t_1 + t_2, 0) \\ - g(t_1 - t_2, 0)] - 2\Sigma(-1 | \tau_3) [g(4t_3, 2t_4) - g(4t_3, -2t_4)] = \Sigma g(0, 2t).$$

* *Messenger of Mathematics*, vol. 54 (1924), pp. 166-176.

† *Transactions of the American Mathematical Society*, vol. 22 (1921), pp. 198-219.

$$(I_2) \quad 4n = m_1 + m_2, \quad 2n = m_3 + m_4:$$

$$\begin{aligned} & 2\Sigma(-1 | \tau_2) [g(t_1 + t_2, \tau_1 - \tau_2) + g(t_1 - t_2, \tau_1 + \tau_2) - g(t_1 + t_2, 0) \\ & \quad - g(t_1 - t_2, 0)] - 2\Sigma(-1 | \tau_3) [g(2t_3, 2t_4) - g(2t_3, -2t_4)] = 0. \end{aligned}$$

$$(II) \quad m = m_1 + 2n_2:$$

$$\begin{aligned} & 2\Sigma[f(t_1 + 2t_2, \tau_1 - \tau_2) - f(t_1 - 2t_2, \tau_1 + \tau_2) \\ & \quad - (-1 | \tau_1 \tau_2) \{f(t_1 + 2t_2, 0) + f(t_1 - 2t_2, 0)\} \\ & \quad - (-1)^{\delta_2} \{f(t_1, -2d_2) - f(t_1, 2d_2)\}] \\ & = \Sigma[\{(-1 | \tau) - 1\}f(t, 0) - 2 \sum_{r=1}^{(\tau-1)/2} f(t, 2r)]. \end{aligned}$$

In this we have the first instance of one of the variables in the partition, here n_2 , being separated into pairs of conjugate divisors of different types, namely, $n_2 = t_2 \tau_2$ and $n_2 = d_2 \delta_2$. The second type can be reduced to the first, but the above statement is the simpler. Similarly in several subsequent identities.

$$(III) \quad m = m_1 + 2n_2:$$

$$\begin{aligned} & 2\Sigma(-1 | \tau_2) [g(t_1 + 2t_2, \tau_1 - \tau_2) + g(t_1 - 2t_2, \tau_1 + \tau_2) - g(t_1 + 2t_2, 0) \\ & \quad - g(t_1 - 2t_2, 0)] + 2\Sigma(-1 | \tau_1) (-1)^{n_2 + d_2 + \delta_2} [g(t_1, 2d_2) - g(t_1, -2d_2)] \\ & = \Sigma[\{1 - (-1\tau)\}g(t, 0) - 2(-1 | \tau) \sum_{r=1}^{(\tau-1)/2} (-1)^r g(t, 2r)]. \end{aligned}$$

$$(IV) \quad m = m_1 + 2n_2 = m_3 + 4n_4:$$

$$\begin{aligned} & 2\Sigma[(-1)^{n_2} f(t_1 - 2t_2, \tau_1 - \tau_2) - f(t_1 - 2t_2, \tau_1 + \tau_2) \\ & \quad - (-1 | t_1 \tau_2) \{f(t_1, 2t_2) + f(t_1, -2t_2)\}] \\ & \quad + 2\Sigma(-1)^{\delta_4} [f(t_3, 2d_4) - f(t_3, -2d_4)] \\ & = \Sigma[\{(-1 | t) - 1\}f(t, 0) - 2 \sum_{r=1}^{(\tau-1)/2} f(t, 2r)]. \end{aligned}$$

$$(V) \quad m = m_1 + 2n_2 = m_3 + 4n_4:$$

$$\begin{aligned} & 2\Sigma(-1 | \tau_2) [(-1)^{n_2} \{g(t_1 + 2t_2, \tau_1 - \tau_2) + g(t_1 - 2t_2, \tau_1 + \tau_2)\} \\ & \quad - (-1 | m_1) \{g(t_1, 2t_2) + g(t_1, -2t_2)\}] \\ & \quad + 2\Sigma(-1 | \tau_3) (-1)^{\delta_4 + \delta_4} [g(t_3, 2d_4) - g(t_3, -2d_4)] \\ & = \Sigma[\{(-1 | m) - (-1 | \tau)\}g(t, 0) - 2(-1 | \tau) \sum_{r=1}^{(\tau-1)/2} (-1)^r g(t, 2r)]. \end{aligned}$$

$$(VI) \quad m = m_1 + 2n_2 = m_3 + 4n_4:$$

$$\begin{aligned} & \Sigma[(-1 | \tau_1 \tau_2) \{f(t_1, \tau_2) + f(t_1, -\tau_2)\} + (-1)^{n_2} \{f(\tau_2, -t_1) - f(\tau_2, t_1)\}] \\ & \quad + \Sigma(-1)^{\delta_4} [f(t_3 + 2d_4, \tau_3 - 2\delta_4) - f(t_3 - 2d_4, \tau_3 + 2\delta_4)] \\ & = \Sigma \sum_{r=1}^{(\tau-1)/2} [f(2r - 1, t) + (-1 | \tau) (-1)^r f(t, 2r - 1)]. \end{aligned}$$

$$(VII) \quad n = n_1 + n_2 = n_3 + 2n_4; \quad 2n = m_5 + m_6:$$

$$\begin{aligned} & \Sigma(-1 | \tau_2) [g(2t_1 + 2t_2, \tau_1 - \tau_2) + g(2t_1 - 2t_2, \tau_1 + \tau_2)] \\ & \quad - 2\Sigma(-1 | \tau_3) [g(2t_3, 2\tau_4) - g(2t_3, -2\tau_4)] \end{aligned}$$

$$\begin{aligned}
& -\Sigma(-1|\tau_6)[g(t_6+t_6,0)+g(t_6-t_6,0)] \\
& = \frac{1}{2}\{1+(-1)^n\}\Sigma g(0,2\tau)-\Sigma(-1|\tau)[g(2t,0) \\
& \quad + \sum_{r=1}^{(\tau-1)/2}\{(-1)^r g(2t,2r)+g(2t,-2r)\}].
\end{aligned}$$

$$\begin{aligned}
\text{(VIII)} \quad n &= n_1 + n_2 = m_3 + 2n_4: \\
2\Sigma(-1)^{n_2}[f(t_1+t_2, \tau_1-\tau_2)-f(t_1-t_2, \tau_1+\tau_2)+f(0, \tau_1+\tau_2)-f(0, \tau_1-\tau_2)] \\
& + 4\Sigma(-1)^{\delta_4}[f(\tau_3, 2d_4)-f(\tau_3, -2d_4)] \\
& = \{1+(-1)^n\}[\zeta'_0(n)f(0,0)-\Sigma f(t,0)] \\
& - 2\Sigma \sum_{r=1}^{(\tau-1)/2} [f(t,2r)+(-1)^n f(t,-2r)-\{1+(-1)^n\}f(0,2r)].
\end{aligned}$$

Here $\zeta'_0(n) \equiv$ the number of odd divisors of n .

$$\begin{aligned}
\text{(IX)} \quad n &= n_1 + n_2; \quad 2n = m_3 + m_4: \\
\Sigma[(-1|\tau_2)g(2t_1+\tau_2, \tau_1-2t_2)+g(2t_1-\tau_2, \tau_1+2t_2) \\
& - (-1|\tau_1)(-1)^{n_2}\{g(\tau_1, \tau_2)-g(\tau_1, -\tau_2)\}] \\
& - \Sigma(-1|\tau_4)[g(t_3, t_4)+g(t_3, -t_4)] \\
& = \Sigma \sum_{r=1}^t [(-1|\tau)g(-\tau, 2r-1)-(-1)^{t+r}g(2r-1, \tau)].
\end{aligned}$$

$$\begin{aligned}
\text{(X)} \quad n &= n_1 + n_2; \quad 2n = m_3 + m_4: \\
\Sigma[(-1)^{n_2}\{f(2t_1-\tau_2, \tau_1+2t_2)-f(2t_1+\tau_2, \tau_1-2t_2) \\
& - (-1)^{n_1}f(\tau_1, -\tau_2)-f(\tau_1, \tau_2)\}-\Sigma(-1|\tau_3\tau_4)[f(t_3, \tau_4)+f(t_3, -\tau_4)]] \\
& = \Sigma \sum_{r=1}^t [(-1)^n f(\tau, -2r+1)-f(2r-1, \tau)].
\end{aligned}$$

$$\begin{aligned}
\text{(XI)} \quad m &= m_1 + 2n_2 = n_3 + n_4: \\
\Sigma[f(t_1+d_2, \tau_1-2d_2)-f(t_1-d_2, \tau_1+2d_2)+(-1)^{\delta_2}f(0, \tau_1-2d_2) \\
& - f(0, \tau_1+2d_2)]-\Sigma(-1)^{\delta_3}[f(d_3, \tau_4)-f(d_3, -\tau_4)] \\
& = \Sigma \left[\sum_{r=1}^{t-1} f(r, \tau) - \sum_{r=1}^{(\tau-1)/2} \{f(t, 2r-1)+f(0, 2r-1)\} \right].
\end{aligned}$$

$$\begin{aligned}
\text{(XI)}_2 \quad 2n &= n_1 + n_2; \quad n = n_3 + n_4: \\
\Sigma[(-1)^{\delta_4}\{f(0, \tau_3+2d_4)-f(0, \tau_3-2d_4)+f(2t_3+d_4, \tau_3-2d_4) \\
& - f(2t_3-d_4, \tau_3+2d_4)\}]-\Sigma(-1)^{\delta_1}[f(d_1, \tau_2)-f(d_1, -\tau_2)] \\
& = \Sigma \left[\sum_{r=1}^{\delta} \{(-1)^{\delta} f(0, 2r-1)-f(d, -2r+1)\} + f(0, \tau) + \sum_{r=1}^{2t-1} f(r, \tau) \right. \\
& \quad \left. + \sum_{r=1}^{(\tau-1)/2} \{f(0, 2r-1)-f(2t, 2r-1)\} \right].
\end{aligned}$$

$$\begin{aligned}
\text{(XII)}_1 \quad m &= m_1 + 2n_2 = n_3 + n_4: \\
\Sigma(-1)^{\delta_2}[f(t_1-d_2, \tau_1+2d_2)-f(t_1+d_2, \tau_1-2d_2)+f(d_2, \tau_1)-f(d_2, -\tau_1)] \\
& - \Sigma(-1|\tau_3\tau_4)[f(t_3, \tau_4)+f(t_3, -\tau_4)] \\
& = \Sigma \left[\frac{1}{2}\{(-1|\tau)-1\}f(0, \tau) - \sum_{r=1}^{\tau-1} f(r, t) - (-1|\tau) \sum_{r=1}^{(\tau-1)/2} (-1)^r f(t, 2r-1) \right].
\end{aligned}$$

$$(XII_2) \quad n = n_1 + n_2; \quad 2n = n_3 + n_4:$$

$$\begin{aligned} & \mathfrak{Z}(-1)^{\delta_2} [f(2t_1 - d_2, \tau_1 + 2\delta_2) - f(2t_1 + d_2, \tau_1 - 2\delta_2) + f(d_2, \tau_1) - f(d_2, -\tau_1)] \\ & \quad - \mathfrak{Z}(-1 | \tau_3 \tau_4) [f(t_3, \tau_4) + f(t_3, -\tau_4)] \\ & = \mathfrak{Z}[(-1)^{\delta} \sum_{r=1}^{\delta} f(d, -2r + 1) + \frac{1}{2}\{(-1 | \tau) - 1\}f(0, \tau) - \sum_{r=1}^{2t-1} f(r, \tau) \\ & \quad - (-1 | \tau) \sum_{r=1}^{(\tau-1)/2} (-1)^r f(2t, 2r - 1)]. \end{aligned}$$

$$(XIII_1) \quad m = m_1 + 2n_2 = n_3 + n_4:$$

$$\begin{aligned} & 2\mathfrak{Z}(-1)^{\delta_2 + \delta_3} [f(t_1 + d_2, \tau_1 - 2\delta_2) - f(t_1 - d_2, \tau_1 + 2\delta_2) + f(d_2, \tau_1) - f(d_2, -\tau_1)] \\ & \quad + 2\mathfrak{Z}(-1 | \tau_3 \tau_4) (-1)^{n_4} [f(t_3, \tau_4) + f(t_3, -\tau_4)] \\ & = 2\mathfrak{Z}[(-1 | \tau) \sum_{r=1}^{(\tau-1)/2} (-1)^r f(t, 2r - 1) - \sum_{r=1}^{\tau-1} (-1)^r f(r, t)] \\ & \quad + \mathfrak{Z}[(-1 | \tau) - 1]f(0, \tau). \end{aligned}$$

$$(XIII_2) \quad n = n_1 + n_2, \quad 2n = n_3 + n_4:$$

$$\begin{aligned} & 2\mathfrak{Z}(-1)^{\delta_2 + \delta_3} [f(2t_1 + d_2, \tau_1 - 2\delta_2) - f(2t_1 - d_2, \tau_1 + 2\delta_2) - f(d_2, \tau_1) + f(d_2, -\tau_1)] \\ & \quad + 2\mathfrak{Z}(-1 | \tau_3 \tau_4) (-1)^{n_4} [f(t_3, \tau_4) + f(t_3, -\tau_4)] \\ & = 2\mathfrak{Z}[\sum_{r=1}^{2t-1} (-1)^r f(r, \tau) + (-1 | \tau) \sum_{r=1}^{(\tau-1)/2} (-1)^r f(2t, 2r - 1)] \\ & \quad + \mathfrak{Z}[1 - (-1 | \tau)]f(0, \tau) - 2\mathfrak{Z}(-1)^{\delta + \delta} \sum_{r=1}^{\delta} f(d, -2r + 1). \end{aligned}$$

$$(XIV_1) \quad m = n_1 + n_2 = m_3 + 2n_4:$$

$$\begin{aligned} & \mathfrak{Z}(-1)^{\delta_2 + \delta_3} [f(d_2, \tau_1) - f(d_2, -\tau_1)] \\ & \quad + \mathfrak{Z}(-1)^{\delta_4} [f(t_3 + \delta_4, \tau_3 - 2d_4) - f(t_3 - \delta_4, \tau_3 + 2d_4) \\ & \quad + f(0, \tau_3 + 2d_4) - f(0, \tau_3 - 2d_4)] \\ & = \mathfrak{Z}[\sum_{r=1}^{(\tau-1)/2} \{f(0, 2r - 1) - f(t, 2r - 1)\} - \sum_{r=1}^{\tau-1} (-1)^r f(r, t)]. \end{aligned}$$

$$(XIV_2) \quad 2n = n_1 + n_2, \quad n = n_3 + n_4:$$

$$\begin{aligned} & \mathfrak{Z}(-1)^{\delta_2 + \delta_3} [f(d_2, -\tau_1) - f(d_2, \tau_1)] \\ & \quad + \mathfrak{Z}(-1)^{\delta_4} [f(0, \tau_3 + 2d_4) - f(0, \tau_3 - 2d_4) \\ & \quad + f(2t_3 + \delta_4, \tau_3 - 2d_4) - f(2t_3 - \delta_4, \tau_3 + 2d_4)] \\ & = \mathfrak{Z}[f(0, \tau) + \sum_{r=1}^{(\tau-1)/2} \{f(0, 2r - 1) - f(2t, 2r - 1)\} + \sum_{r=1}^{2t-1} (-1)^r f(r, \tau)] \\ & \quad + \mathfrak{Z}(-1)^{\delta} \sum_{r=1}^{\delta} [f(0, 2r - 1) - f(d, -2r + 1)]. \end{aligned}$$

$$(XV) \quad n = n_1 + n_2, \quad 2n = n_3 + n_4:$$

$$\begin{aligned} & 2\mathfrak{Z}(-1)^{\delta_2} [f(2d_1 - 2d_2, \delta_1 + \delta_2) - f(2d_1 + 2d_2, \delta_1 - \delta_2) \\ & \quad - 2f(2d_2, -\tau_1) + 2f(2d_2, \tau_1)] - 2\mathfrak{Z}(-1)^{n_4} [f(\tau_3 - \tau_4, 0) - f(\tau_3 + \tau_4, 0)] \\ & = 2\mathfrak{Z}[f(0, 0) + 2 \sum_{r=1}^{(\tau-1)/2} f(2r, 0)] + \mathfrak{Z}[2(-1)^{\delta} f(2d, 0) - \{1 + (-1)^{\delta}\}f(0, \delta) \\ & \quad - 2 \sum_{r=1}^{\delta-1} \{f(2r, d) + (-1)^{\delta} f(2r, -d) - (-1)^{\delta} f(2d, -r) - (-1)^{\delta+r} f(2d, r)\}]. \end{aligned}$$

$$(XVI) \quad 2n = m_1 + m_2:$$

$$\Sigma[f(t_1 - t_2, \tau_1 + \tau_2) - f(t_1 + t_2, \tau_1 - \tau_2)] = \Sigma t[f(0, 2t) - f(2t, 0)].$$

$$(XVII_1) \quad 2m = m_1 + m_2, \quad m = n_3 + n_4:$$

$$\begin{aligned} & \Sigma(-1 | m_2)[f(t_1 - t_2, \tau_1 + \tau_2) - f(t_1 + t_2, \tau_1 - \tau_2)] \\ & + \Sigma(-1)^{n_3}(-1 | \tau_3 \tau_4)[f(2t_3, 2t_4) - f(2t_3, 2t_4) + f(2t_4, -2t_3) - f(2t_3, -2t_4)] \\ & = \Sigma(-1 | \tau)[f(0, 2t) - f(2t, 0)]. \end{aligned}$$

$$(XVII_2) \quad 4n = m_1 + m_2, \quad 2n = n_3 + n_4:$$

$$\begin{aligned} & \Sigma(-1 | m_2)[f(t_1 - t_2, \tau_1 + \tau_2) - f(t_1 + t_2, \tau_1 - \tau_2)] \\ & = \Sigma(-1)^{n_3}(-1 | \tau_3 \tau_4)[f(2t_3, 2t_4) - f(2t_4, 2t_3) + f(2t_3, -2t_4) - f(2t_4, -2t_3)] \end{aligned}$$

$$(XVIII) \quad m = m_1 + 4n_2 = m_3 + 2n_4:$$

$$\begin{aligned} & \Sigma[f(t_1 - 2d_2, \tau_1 + 2\delta_2) - f(t_1 + 2d_2, \tau_1 - 2\delta_2)] \\ & + \Sigma[f(t_3, -\tau_4) - f(t_3, \tau_4) + f(\tau_4, t_3) - f(\tau_4, -t_3)] \\ & = \sum_{r=1}^{(\tau-1)/2} [f(t, 2r-1) - f(2r-1, t)]. \end{aligned}$$

$$(XIX) \quad m = m_1 + 4n_2 = m_3 + 2n_4:$$

$$\begin{aligned} & \Sigma(-1)^{d_2 + \delta_2}[f(t_1 + 2d_2, \tau_1 - 2\delta_2) - f(t_1 - 2d_2, \tau_1 + 2\delta_2)] \\ & + \Sigma(-1 | \tau_3 \tau_4)(-1)^{n_4}[f(t_3, \tau_4) + f(t_3, -\tau_4) - f(\tau_4, t_3) - f(\tau_4, -t_3)] \\ & = \Sigma(-1 | \tau) \sum_{r=1}^{(\tau-1)/2} (-1)^r [f(t, 2r-1) - f(2r-1, t)]. \end{aligned}$$

$$(XX_1) \quad m = n_1 + n_2:$$

$$\begin{aligned} & 2\Sigma[f(t_1 + t_2, \tau_1 - \tau_2) - f(t_1 - t_2, \tau_1 + \tau_2)] \\ & = \Sigma[(t-1)f(t, 0) - \sum_{r=1}^{(\tau-1)/2} \{f(t, 2r) + f(t, -2r)\}]. \end{aligned}$$

$$(XX_2) \quad 2n = n_1 + n_2:$$

$$\begin{aligned} & 2\Sigma[f(t_1 + t_2, \tau_1 - \tau_2) - f(t_1 - t_2, \tau_1 + \tau_2)] \\ & = \Sigma[(2t-1)f(2t, 0) - \sum_{r=1}^{(\tau-1)/2} \{f(2t, 2r) + f(2t, -2r)\} - df(0, 2d)] \end{aligned}$$

$$(XXI_1) \quad m = n_1 + n_2 = m_3 + 2n_4:$$

$$\begin{aligned} & \Sigma(-1)^{n_2}(-1 | \tau_2)[g(t_1 + t_2, \tau_1 - \tau_2) + g(t_1 - t_2, \tau_1 + \tau_2)] \\ & + \Sigma(-1)^{\delta_4}(-1 | \tau_3)\{1 + (-1)^{d_4}\}[g(t_3, 2d_4) - g(t_3, -2d_4)] \\ & = \Sigma(-1 | \tau) \sum_{r=1}^{(\tau-1)/2} [g(t, -2r) - (-1)^r g(t, 2r)]. \end{aligned}$$

$$(XXI_2) \quad 2n = n_1 + n_2, \quad n = n_3 + n_4:$$

$$\begin{aligned} & 2\Sigma(-1)^{n_2}(-1 | \tau_2)[g(t_1 + t_2, \tau_1 - \tau_2) + g(t_1 - t_2, \tau_1 + \tau_2)] \\ & - 2\Sigma(-1)^{\delta_4}(-1 | \tau_3)\{1 - (-1)^{d_4}\}[g(2t_3, 2d_4) - g(2t_3, -2d_4)] \\ & = -2\Sigma(-1 | \tau)[g(2t, 0) + \sum_{r=1}^{(\tau-1)/2} \{(-1)^r g(2t, 2r) + g(2t, -2r)\}] \\ & + \Sigma(-1)^{\delta}\{1 - (-1)^d\}g(0, 2d). \end{aligned}$$

$$(XXII) \quad 2n = n_1 + n_2 = m_3 + m_4:$$

$$\begin{aligned} & \Sigma[f(2t_1 + \tau_2, \tau_1 - 2t_2) - f(2t_1 - \tau_2, \tau_1 + 2t_2) + f(\tau_1, -\tau_2) - f(\tau_1, \tau_2)] \\ & \quad + \Sigma[f(t_3, t_4) - f(t_3, -t_4)] \\ & = \Sigma\left[\sum_{r=1}^t \{f(2r-1, \tau) - f(\tau, -2r+1)\}\right]. \end{aligned}$$

$$(XXIII) \quad n = n_1 + n_2, \quad 2n = m_3 + m_4:$$

$$\begin{aligned} & \Sigma(-1 \mid \tau_1)[(-1)^{n_1}\{g(2t_2 + \tau_1, \tau_2 - 2t_1) + g(2t_2 - \tau_1, \tau_2 + 2t_1) \\ & \quad - (-1)^n\{g(\tau_1, \tau_2) - g(\tau_1, -\tau_2)\}\}] \\ & \quad - \Sigma(-1 \mid \tau_3 m_4)[g(t_3, t_4) - g(t_3, -t_4)] \\ & = -(-1)^n \sum_{r=1}^t [(-1 \mid \tau)g(\tau, -2r+1) + (-1)^r g(2r-1, \tau)]. \end{aligned}$$

$$(XXIV) \quad n = n_1 + n_2:$$

$$\begin{aligned} & \Sigma[f(d_1 - d_2, \delta_1 + \delta_2) - f(d_1 + d_2, \delta_1 - \delta_2)] \\ & = \Sigma[(d-1)\{f(0, d) - f(d, 0)\} \\ & \quad + \sum_{r=1}^{\delta-1} \{f(d, r) - f(r, d) + f(d, -r) - f(r, -d)\}]. \end{aligned}$$

$$(XXV) \quad n = n_1 + n_2:$$

$$\begin{aligned} & 2\Sigma(-1)^{d_2+\delta_2}[f(d_1 + d_2, \delta_1 - \delta_2) - f(d_1 - d_2, \delta_1 + \delta_2) \\ & \quad + (-1)^{\delta_1}\{f(d_1, d_2) - f(d_1, -d_2) - f(d_2, d_1) + f(d_2, -d_1)\}] \\ & = \Sigma(-1)^\delta[1 + (-1)^d]\{f(0, d) - f(d, 0)\} \\ & \quad + 2 \sum_{r=1}^{\delta-1} \{(-1)^{df}(r, -d) - (-1)^{df}(d, -r) \\ & \quad + (-1)^r f(r, d) - (-1)^r f(d, r)\}. \end{aligned}$$

From this set many more can be written down, by elimination of a particular partition, etc., but the set as given is probably in the simplest form.

5. *General Identities.* We shall not take space here to write these out, but will reserve them for another occasion. It will be noticed that the arguments of f or of g in several pairs of the identities in § 4 are the same, and that one identity in particular pairs of this kind involves f , the other, g . Hence each such pair is equivalent to a single identity involving the function $h(x, y)$, which is finite and single-valued when x, y are simultaneously integers, and which otherwise is *completely arbitrary*. For, we may write

$$h(x, y) \equiv \frac{1}{2}[h(x, y) + h(-x, -y)] + \frac{1}{2}[h(x, y) - h(-x, -y)],$$

and the first [] is an *instance* of $f(x, y)$, the second, of $g(x, y)$. These identities involving h are applicable to certain arithmetical forms of arbitrary degree.

DETERMINATION OF THE GROUPS OF ORDERS 162-215 OMITTING ORDER 192.

By J. K. SENIOR and A. C. LUNN.

The groups of order g where $100 < g < 162$ and $g \neq 128$ have recently been listed,* and it is a comparatively easy matter to treat the cases where $161 < g < 216$ and $g \neq 192$. The present paper is therefore a continuation of the one just cited, and the methods and symbolism used are the same as those therein defined.

Between 161 and 216 there are only 7 integers which are the product of more than four prime factors. These are

$$\begin{array}{lll} 162 = 2 \cdot 3^4 & 180 = 2^2 \cdot 3^2 \cdot 5 & 208 = 2^4 \cdot 13 \\ 168 = 2^3 \cdot 3 \cdot 7 & 192 = 2^6 \cdot 3 & \\ 176 = 2^4 \cdot 11 & 200 = 2^3 \cdot 5^2 & \end{array}$$

The groups of order 168 have been listed by G. A. Miller †; those of orders 176 and 208 by Lunn and Senior.‡ To determine the number groups of order 192 is very laborious, and no attempt is made here to solve the problem. But brief arguments suffice to cover the orders 162, 180 and 200 which are here treated in some detail. For the orders where g is the product of less than five factors, since the general methods are known, only the results are given.

THE GROUPS OF ORDER $162 = 2 \cdot 3^4$.

Every group of order 162 is solvable and thus determines a $(G_{81}^{81} : G_2^2)_k$, ($k = 1$ or 2). Hence every group of order 162 occurs in one of the following divisions:

$$\text{Division (a)} \quad (G_{81}^{81} : G_2^2)_1 \qquad \text{Division (b)} \quad (G_{162}^{81} : G_2^2)_2.$$

Division (a). $(G_{81}^{81} : G_2^2)_1$. A group in this division is the direct product of its Sylow subgroups. Since there are fifteen groups of order 81, and one group of order 2, there are fifteen groups of division (a). Five of these are abelian.

Division (b). $(G_{162}^{81} : G_2^2)_2$. A group in this division corresponds to a set of conjugate subgroups of order 2 in the i -group of a group of order 81. The fifteen groups of this latter order are therefore considered one at a time. In the case of each i -group, the number of sets of conjugate subgroups of order 2 has been proven by the authors, but, in order not to expand the treatment unduly, the proofs are here omitted and only the results given. In the following table, each group of order 81 or 162 is defined by the relations of its generators, which are labelled A-E.

* Senior and Lunn, *American Journal of Mathematics*, vol. 56 (1934), p. 328.

† G. A. Miller, *American Mathematical Monthly*, vol. 9 (1902), p. 1.

‡ Lunn and Senior, *American Journal of Mathematics*, vol. 56 (1934), p. 321.

TABLE I.

GROUPS OF ORDER 81

GROUPS OF ORDER 162

	$B^{-1}AB$	$C^{-1}AC$	$C^{-1}BC$	$D^{-1}AD$	$D^{-1}BD$	$D^{-1}CD$	$E^{-1}AE$	$E^{-1}BE$	$E^{-1}CE$	$E^{-1}DE$	$E^2 = 1$
1	$A^{81} = 1$						1	A^{-1}			
2	$A^{27} = B^3 = 1$	A					2	A^{-1}	B		
							3	A	B^{-1}		
							4	A^{-1}	B^{-1}		
3	$A^{27} = B^3 = 1$	A^{10}					5	A^{-1}	B		
4	$A^9 = B^9 = 1$	A					6	A	B^{-1}		
							7	A^{-1}	B^{-1}		
5	$A^9 = B^3 = C^3 = 1$	A	A	B			8	A	B	C^{-1}	
							9	A	B^{-1}	C^{-1}	
							10	A^{-1}	B	C	
							11	A^{-1}	B	C^{-1}	
							12	A^{-1}	B^{-1}	C^{-1}	
6	$A^9 = B^3 = 1$	$C^3 = A^3$	A	A^4B	BA^6		13	A^{-1}	B	C^{-1}	
7	$A^9 = B^3 = C^3 = 1$	A	A^4B	BA^6			14	A^{-1}	B^{-1}	C	
							15	A	B^{-1}	C^{-1}	
							16	A^{-1}	B	C^{-1}	
8	$A^9 = B^3 = C^3 = 1$	A	A^7B	BA^3			17	A^{-1}	B^{-1}	C	
							18	A	B^{-1}	C^{-1}	
							19	A^{-1}	B	C^{-1}	
9	$A^9 = B^3 = C^3 = 1$	A	A^4	B			20	A	B^{-1}	C	
							21	A^{-1}	B	C	
							22	A^{-1}	B^{-1}	C	
10	$A^9 = B^3 = C^3 = 1$	A	A	BA^3			23	A	B^{-1}	C^{-1}	
							24	A^{-1}	B^{-1}	C	
11	$A^9 = B^9 = 1$	A^4					25	A^{-1}	B		
12	$A^9 = B^3 = C^3 = 1$	A	AB	B			26	A	B^{-1}	C^{-1}	
							27	A^{-1}	B	C^{-1}	
							28	A^{-1}	B^{-1}	C	
13	$A^3 = B^3 = C^3 = D^3 = 1$	A	A	B	A	B	29	A	B	C	D^{-1}
							30	A	B	C^{-1}	D^{-1}
							31	A	B^{-1}	C^{-1}	D^{-1}
							32	A^{-1}	B^{-1}	C^{-1}	D^{-1}
14	$A^3 = B^3 = C^3 = D^3 = 1$	A	A	BA	A	B	33	A	B	C	D^{-1}
							34	A	B^{-1}	C^{-1}	D
							35	A^{-1}	B	C^{-1}	D
							36	A	B^{-1}	C^{-1}	D^{-1}
							37	A^{-1}	B	C^{-1}	D^{-1}
15	$A^3 = B^3 = C^3 = D^3 = 1$	A	A	B	A	CBA^{-1}	38	A^{-1}	B^{-1}	C^{-1}	D
							39	A	B^{-1}	C	D^{-1}
							40	A^{-1}	B	C^{-1}	D^{-1}

The number of groups of order 162 is thus:

Division (a)	Division (b)
15	40

THE GROUPS OF ORDER $180 = 2^2 \cdot 3^2$

It is well known that there is just one insolvable A solvable group of this order determines a $(G^4_{4k_1} : G^9_{9k_2} : ($
 $k_1 = 1$ or 3 , since 4 and 12 are the only orders which transitive groups of degree four exist.

$k_2 = 1, 2$, or 4 , since 9, 18 and 36 are the only ord for which transitive groups of degree nine exist.

$k_3 = 1, 2$, or 4 , since 5, 10 and 20 are the only ord transitive groups of degree five exist.

Thus every solvable group of order 180 occurs in divisions:

Division (a)	$(G^4_4 : G^9_9 : G^5_5)_1$	Division (f)	$[(G^4_4 : G^9_9 : G^5_5)_1 : G^2_2]_1$
" (b)	$[(G^4_4 : G^9_{18})_2 : G^5_5]_1$	" (g)	$[(G^4_4 : G^9_{18})_2 : G^5_5]_1$
" (c)	$[(G^4_4 : G^5_{10})_2 : G^9_9]_1$	" (h)	$[(G^4_4 : G^5_{10})_2 : G^9_9]_1$
" (d')	$[(G^9_{18} : G^5_{10})_2 : G^4_4]_2$	" (i)	$[(G^9_{18} : G^5_{10})_2 : G^4_4]_2$
" (d'')	$[(G^9_{18} : G^5_{10})_1 : G^4_4]_4$	" (j)	$[(G^9_{18} : G^5_{10})_1 : G^4_4]_4$
" (e)	$[(G^5_{20} : G^4_4)_4 : G^9_9]_1$		

Division (a). $(G^4_4 : G^9_9 : G^5_5)_1$. A group in this product of its Sylow subgroups. There are two groups order nine, and one of order five. Hence there are 2 division (a). They are all abelian.

Division (b). $[(G^4_4 : G^9_{18})_2 : G^5_5]_1$. Each of the 2 dimidiated in one way with each of the three groups \bar{G} $2 \times 3 = 6$ groups of division (b).

Division (c). $[(G^4_4 : G^5_{10})_2 : G^9_9]_1$. Each of the tw can be multiplied directly by each of the two groups G $2 \times 2 = 4$ groups of division (c).

Division (d'). $[(G^9_{18} : G^5_{10})_2 : G^4_4]_2$. There are three Each of these can be dimidiated in one way with each of Hence there are $3 \times 2 = 6$ groups of division (d').

Division (d''). $[(G_{18}^9:G_{10}^5)_1:G_4^4]_4$. There are three groups G_{18}^9 and one group G_{10}^5 . Hence there are three groups of division (d'').

Divisions (e), (f) and (g). $[(G_{20}^5:G_4^4)_4:G_9^9]_1$, $[(G_{20}^5:G_4^4)_4:G_{18}^9]_2$ and $[(G_{20}^5:G_4^4)_4:G_{36}^9]_4$. There is only one group $(G_{20}^5:G_4^4)_4$. Hence there are two groups of division (e). This group of order 20 can be dimidiated in only one way, and hence yields with the three groups G_{18}^9 the three groups of division (f). The only quotient group of order four in this group of order 20 is cyclic. As there is only one case of such a quotient group among the two groups G_{36}^9 , and as this quotient group gives rise to only a single isomorphism, there is one group of division (g).

Divisions (h) and (i). $[(G_{36}^9:G_4^4)_4:G_5^5]_1$ and $[(G_{36}^9:G_4^4)_4:G_{10}^5]_2$. There are two groups $(G_{36}^9:G_4^4)_4$ which permit in all three distinct dimidiations. Thus with the one group G_5^5 they yield the two groups of division (h), and with the one group G_{10}^5 , they yield the three groups of division (i).

Division (j). $[(G_{12}^4:G_{9k_2}^9)_{3k_2}:G_{5k_3}^5]_{k_2}$. There is only one group G_{12}^4 . It contains no invariant subgroup of order two and hence $k_2 \neq 2$. Neither group G_{36}^9 contains a quotient group simply isomorphic with G_{12}^4 and so $k_2 \neq 4$. Thus $k_2 = 1$ and a group of division (j) is $[(G_{12}^4:G_9^9)_3:G_{5k_3}^5]_{k_3}$. As neither of the two groups $(G_{12}^4:G_9^9)_3$ contains an invariant subgroup of index 2 or 4, $k_3 \neq 2$ or 4. Hence $k_3 = 1$ and there are two groups of division (j).

The number of groups of order 180 is thus:

Insolvable	1	
Solvable	Division (a)	4
“	“ (b)	6
“	“ (c)	4
“	“ (d')	6
“	“ (d'')	3
“	“ (e)	2
“	“ (f)	3
“	“ (g)	1
“	“ (h)	2
“	“ (i)	3
“	“ (j)	2
Total		37

THE GROUPS OF ORDER $200 = 2^3 \cdot 5^2$.

Every group of order 200 is solvable and thus determines a $(G_{8k_1}^8 : G_{25k_2}^{25})_{k_1k_2}$. $k_1 = 1$; as 8 is the only order which divides 200 for which transitive groups of degree 8 exist. $k_2 = 1, 2, 4$ or 8. Every group of order 200 occurs therefore in one of the following divisions:

Division (a) $(G_8^8 : G_{25}^{25})_1$

Division (c) $(G_{100}^8 : G_{100}^{25})_4$

Division (b) $(G_8^8 : G_{50}^{25})_2$

Division (d) $(G_8^8 : G_{200}^{25})_8$

Division (a). $(G_8^8 : G_{25}^{25})$. A group in this division is the direct product of its Sylow subgroups. As there are five groups of order 8, and two groups of order 25, there are $5 \times 2 = 10$ groups of division (a). Six of these are abelian.

Division (b). $(G_8^8 : G_{50}^{25})$. The five groups of order 8 permit seven distinct dimidiations; the three groups G_{50}^{25} permit one dimidiation each. Hence there are $7 \times 3 = 21$ groups of division (b).

Division (c). $(G_8^8 : G_{100}^{25})$. There are six groups G_{100}^{25} . Five of them involve one case of cyclic quotient group of order four each: the sixth involves one case of non-cyclic quotient group of this order. The groups of division (c) may therefore be divided into two subdivisions.

(1) Quotient group of order 4 cyclic. The five groups of order 8 involve in all two cases of cyclic quotient group of order 4, and each case gives rise to only one isomorphism. Combination with the five groups G_{100}^{25} which involve cyclic quotient groups of order 4 therefore yields $2 \times 5 = 10$ groups of subdivision (1).

(2) Quotient group of order 4 non-cyclic. The one group G_{100}^{25} which involves a non-cyclic quotient group of order 4 contains just one characteristic subgroup of index 2. Hence there arise the following groups of subdivision (2).

Group of order 8	No. of groups of order 200
Cyclic	0
Abelian, type 2, 1.....	2
Dihedral	2
Dicyclic	1
Abelian, type 1, 1, 1.....	1
Total	6

Thus there are $10 + 6 = 16$ groups of division (c).

Division (d). ($G^8_8: G^{25}_{200}$). A group of this division corresponds to a set of conjugate subgroups of order 8 in the i -group of a group of order 25. The i -group of the cyclic group of order 25 contains Sylow subgroups of order 4 and hence gives rise to no groups of this division. The i -group of the non-cyclic group of order 25 contains Sylow subgroups of order 32. The subgroups of order 8 are permuted under the i -group in five sets of conjugates, and thus there arise the five groups of division (d).

The number of groups of order 200 is thus:

Division	(a)	(b)	(c)	(d)	
Number	10	21	16	5	Total 52

There follows a list of the number of groups of every order (except 192) between 161 and 216 where this number exceeds one.

Order	Factors	Number of groups
162	pq^4	55
164	p^2q	5
165	pqr	2
166	pq	2
168	p^3qr	57
169	p^2	2
170	pqr	4
171	p^2q	5
172	p^2q	4
174	pqr	4
175	p^2q	2
176	p^4q	42
178	pq	2
180	p^2q^2r	37
182	pqr	4
183	pq	2
184	p^3q	12
186	pqr	6
188	p^2q	4
189	p^3q	13
190	pqr	4
192	p^6q	not determined
194	pq	2

Order	Factors	Number of groups
195	pqr	2
196	p^2q^2	12
198	pq^2r	10
200	p^3q^2	52
201	pq	2
202	pq	2
203	pq	2
204	p^2qr	12
205	pq	2
206	pq	2
207	p^2q	2
208	p^4q	51
210	$pqrs$	12
212	p^2q	5
214	pq	2

THE UNIVERSITY OF CHICAGO.

A DETERMINATION OF ALL POSSIBLE SYSTEMS OF STRICT IMPLICATION.

By MORGAN WARD.

1°. It is known that the postulates chosen by C. I. Lewis for his "system of strict implication" † are not categorical, since three distinct types of such a system have been shown to exist.‡ I shall prove here that the three types already discovered are the only ones possible. The inclusion of an additional modal postulate ‡ will therefore make the system categorical, and allow it to be exhibited as a four-valued truth-value system. The corresponding entscheidung problem may then be solved by the matrix method.

2°. In what follows, the decimal numeration 11.01-20.01 refers to *Symbolic Logic*, Chapter VI. We shall modify Lewis' notation as follows. We use + instead of \vee to denote logical addition, p' for $\sim p$ and p^* for $\sim \Diamond p$. We shall refer to the system of strict implication as (the system) Σ .

TABLE I.
The System Σ .

Primitive Ideas	Postulates
$p, p', \Diamond p, pq, p = q.$	11.1 $pq \cdot \prec \cdot qp$
	11.2 $pq \cdot \prec \cdot p$
	11.3 $p \cdot \prec \cdot pp$
Definitions	11.4 $(pq)r \cdot \prec \cdot p(qr)$
11.01 $p + q = \cdot (p'q')'.$	11.5 $p \cdot \prec \cdot (p')'$
11.02 $p \prec q = \cdot (pq)^*.$	11.6 $p \prec q \cdot q \prec r : \prec \cdot p \prec r$
11.03 $p = q = : p \prec q \cdot q \prec p$	11.7 $p \cdot p \prec q : \prec q$
	19.01 $\Diamond pq \cdot \prec \cdot \Diamond p$
	20.01 $(\exists p, q) : (p \prec q)' \cdot (p \prec q')'.$

It is also assumed that the system is closed with respect to the unary operations $p' \Diamond p$ and the binary operation pq . The equality relation = of the primitive ideas has the usual properties.§ In the present abstract treatment, 11.03 may be looked upon as a condition upon the relation \prec .

† It is assumed that the reader is familiar with the contents of Chapters VI and VII of C. I. Lewis and C. H. Langford's book, *Symbolic Logic* (New York, 1932), where a detailed account is given both of the system of strict implication and the matrix method as applied to truth-value systems. We shall refer to this book as *Symbolic Logic*.

‡ *Symbolic Logic*, Appendix II.

§ As given, for example, in E. V. Huntington's paper, "Postulates for the algebra of logic," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 279-280.

3°. THEOREM.† The system Σ is a Boolean algebra in which $p + q$ and pq are the operations of addition and multiplication, and p' is the negation of p .

The following set of postulates for a Boolean algebra is given by Huntington in his Transactions paper, page 280. We presuppose a class K of elements p, q, r, \dots a unary operation p' , a binary operation $+$ and an equality relation $=$ which we identify with the corresponding entities of Σ

H_0 [20. 1, 20. 11] K contains at least two distinct elements.

H_2 [11. 01] If p and q are in the class K , then $p + q$ is in the class K .

H_2 [13. 11] $p + q = q + p$.

H_3 [13. 4] $(p + q) + r = p + (q + r)$.

H_4 [13. 31] $p + p = p$.

H_5 [18. 2] $(p' + q')' + (p' + q)' = p$.

Def. H_6 [11. 01, 12. 3] $pq = (p'q')'$.

The numbers in square brackets refer to the corresponding theorems in *Symbolic Logic*.

4°. THEOREM. If the system of strict implication is interpreted as a truth-value system with a finite number of truth-values n_1, n_2, \dots, n_k , then n_1, n_2, \dots, n_k must form a Boolean algebra \mathfrak{B} with respect to the operations of addition, multiplication and negation derived from the matrices for $p + q, pq$ and p' .

For suppose that the matrices for p' and $p + q$ are

p	p'	p	q	n_1	n_2	\dots	n_k
n_1	β_1	n_1		α_{11}	α_{12}	\dots	α_{1k}
n_2	β_2	n_2		α_{21}	α_{22}	\dots	α_{2k}
\cdot	\cdot	\cdot		\cdot	\cdot	\dots	\cdot
\cdot	\cdot	\cdot		\cdot	\cdot	\dots	\cdot
\cdot	\cdot	\cdot		\cdot	\cdot	\dots	\cdot
n_k	β_k	n_k		α_{k1}	α_{k2}	\dots	α_{kk}

where each α and β stands for a definite truth-value n . We then define the operations of negation and addition over n_1, n_2, \dots, n_k by

$$n'_i = \beta_i, \quad n_i + n_j = \alpha_{ij} \quad (i, j = 1, \dots, k)$$

and it is immediately obvious that the conditions $H_0 - H_6$ of section 3° are all satisfied.

† For a detailed analysis of the correspondence between Σ and a Boolean algebra, see E. V. Huntington, *Bulletin of the American Mathematical Society*, vol. 40 (October, 1934), pp. 729-735.

COROLLARY. *The number of truth-values in any representation of Σ as a truth-value system is either infinite or a power of 2.*

Let us use the letters ∂ and ϵ to stand for designated values \dagger and undesignated values in \mathfrak{B} respectively. Then ∂ and ϵ combine in \mathfrak{B} as follows:

TABLE II.
Combination of Truth-Values.

$+$	∂	ϵ	\times	∂	ϵ	$'$	∂	ϵ
ϵ	∂	∂	∂	∂	ϵ	∂	∂	ϵ
∂	∂	ϵ	ϵ	ϵ	ϵ	ϵ	∂	∂

For example, the second table tells us that the product of two designated values is a designated value, the product of a designated value and an undesignated value is an undesignated value, and so on.

These facts result from the obvious propositions of Σ

$$p \cdot q : \prec : p + q \cdot pq; \quad pq' : \prec : p + q' \cdot (p'q)'; \quad p \cdot \prec : (p')'.$$

5°. We consider now the possible representations of Σ as a four-valued truth-value system. In accordance with the results of section 4°, we may take for the set of truth-values \mathfrak{B} the four numbers 1, 2, 3 and 6, which form a Boolean algebra if addition and multiplication are taken as the operations of finding the greatest common divisor and least common multiple, while negation is defined by $1' = 6$, $2' = 3$.

TABLE III.
Truth-Values of p' , p^* and so on.

p	p'	p^*	$p + p'$	pp'	pp'^*	$\diamond p$
1	6	a	1	6	d	a'
2	3	b	1	6	d	b'
3	2	c	1	6	d	c'
6	1	d	1	6	d	d'

There are in all $4^4 = 256$ such interpretations of Σ conceivable obtained by giving each of a, b, c, d , its four possible values 1, 2, 3, or 6. We shall use the definitions and postulates of Σ in Table I to reduce this number to eight.

From Table III, we see that \dagger

- (i) $d = \partial$, (ii) $6 \neq \partial$, (iii) $1 = \partial$.

\dagger *Symbolic Logic*, pp. 231-233.

\ddagger We use the letter " ∂ " to stand for some designated value. Thus $6 \neq \partial$ means that 6 is not a designated value, and $ab, ac, ad = \partial$ would mean that ab, ac , and ad are all designated values.

From the last theorem of 4° and (ii) we see that

$$(iv) \quad \text{if } 2 = \theta, 3 \neq \theta; \text{ if } 3 = \theta, 2 \neq \theta.$$

TABLE IV.

Matrices for pq , pq' and so on.

pq	pq'	$p \prec q$	$q \prec p$	$p = q$			
1 2 3 6	6 3 2 1	$d c b a$	$d d d d$	d	$d c$	$a b$	$d a$
2 2 6 6	6 6 2 2	$d d b b$	$c d c d$	$d c$	d	$b c$	$b d$
3 6 3 6	6 3 6 3	$d c d c$	$b b d d$	$a b$	$b c$	d	$d c$
6 6 6 6	6 6 6 6	$d d d d$	$a b c d$	$d a$	$b d$	$d c$	d

Now since equality over Σ is defined as logical equivalence,† $p = q$ when and only when p and q have the same truth-values. Therefore, we infer from the matrix for $p = q$ that $ad, bc, bd, cd \neq \theta$. Hence by (i) and Table II,

$$(v) \quad a, b, c \neq \theta.$$

From (v), (i) and (iii), we see that

$$(vi) \quad a, b, c \neq d$$

$$(vii) \quad a, b, c \neq 1.$$

TABLE V.

The Principle of the syllogism.

p	q	q'	$p \prec q$	$p \cdot p \prec q$	$p \cdot p \prec q : q'$	11. 7. = $(p \cdot p \prec q : q')^*$
1	1	6	d	d	6	d
1	2	3	c	c	$3c$	$(3c)^*$
1	3	2	b	b	$2b$	$(2b)^*$
1	6	1	a	a	a	a^*
2	1	6	d	$2d$	6	d
2	2	3	d	$2d$	6	d
2	3	2	b	$2b$	$2b$	$(2b)^*$
2	6	1	b	$2b$	$2b$	$(2b)^*$
3	1	6	d	$3d$	6	d
3	2	3	c	$3c$	$3c$	$(3c)^*$
3	3	2	d	$3d$	6	d
3	6	1	c	$3c$	$3c$	$(3c)^*$
6	1	6	d	6	6	d
6	2	3	d	6	6	d
6	3	2	d	6	6	d
6	6	1	d	6	6	d

† Lewis and Langford, pp. 123-124.

From the last column of Table V, we see that

$$(viii) \quad a^*, (2b)^*, (3c)^* = \theta.$$

I say that $a = 6$. For by (vii), $a \neq 1$. And if $a = 2$ or 3 , by (viii), $a^* = 2^*$ or $a^* = 3^*$. Hence $a^* = b$ or c , $= \theta$ contradicting (v).

I say that $b = 3$ or $b = 6$. For by (vii), $b \neq 1$. And if $b = 2$, then by (viii), $(2b)^* = 2^* = b = 2 = \theta$ contradicting (v).

Finally, $c = 2$ or $c = 6$. For by (vii), $c \neq 1$. And if $c = 3$, then by (viii) $(3c)^* = 3^* = c = 3 = \theta$ contradicting (v).

We cannot have $b = 3$ and $c = 2$. For then $d = 1$ by (ii) and (v). Hence $\langle \rangle p \cdot = p'$ and Σ will degenerate into a system of material implication, contradicting 20.01.

We summarize our results in the following

THEOREM. *There are at most eight possible four-valued systems of strict implication, distinguished by the truth-values of $\langle \rangle p$; namely*

TABLE VI.
Possible Systems Σ .

p	$\langle \rangle p$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1		1	1	1	1	1	1	1	1
2		1	1	2	2	1	1	1	1
3		3	3	1	1	1	1	1	1
6		6	2	6	3	6	6	3	2
Designated									
Values †		1, 3	1, 3	1, 2	1, 2	1, 2	1, 2	1, 2	1, 3

These systems may be grouped into four pairs, (7) and (8); (1) and (3); (5) and (6); (2) and (4); which are permuted into one another by the interchange of the truth-values 2 and 3, and are hence not essentially distinct. Finally, the four pairs are immediately seen to agree with the systems called Group I, Group II, Group III and Group V, in Appendix II of *Symbolic Logic*.

I have verified that the first three pairs satisfy all the postulates of Σ , while the last pair satisfy all the postulates save 19.01, as was first proved by W. T. Parry, M. Wajsberg and P. Henle.† I shall denote these three systems of strict implication by Σ_1 , Σ_2 , Σ_3 .

† Obtained by (i), (ii) and (iv).

‡ *Symbolic Logic*, footnote, page 492.

6°. It remains to show that there is no representation of Σ as a truth-value system of finite order † essentially distinct from Σ_1 , Σ_2 and Σ_3 .

Suppose that a representation of Σ as a truth-value system maps Σ upon a Boolean algebra \mathfrak{B}_N of order 2^N , $N \geq 3$ such that all the postulates of Σ are satisfied in accordance with the matrix method.

Let N generating elements of the algebra \mathfrak{B}_N be $\alpha_1, \alpha_2, \dots, \alpha_N$. Since $N \geq 3$, we see from Table II that there are at least two generators which are both designated values, or at least two generators which are undesignated values. With a proper choice of notation, we may assume that α_1, α_2 are such a pair.

Now every element ν of the algebra \mathfrak{B}_N may be uniquely represented in the form

$$(1) \quad \nu = \alpha_1^{e_1} \alpha_2^{e_2} \cdots \alpha_N^{e_N}$$

where the exponents e are either zero or one, and by convention, the universal element of \mathfrak{B}_N is denoted by 1, $\alpha^0 = 1$.

Consider now the effect of equating α_1 and α_2 . An inspection of Table II and (1) shows us that this operation does not convert any designated value into an undesignated value, or vice versa. Hence the truth-value table establishing the validity of any one of our postulates for Σ in \mathfrak{B}_N , is unaffected by the operation.‡

This operation, however, throws \mathfrak{B}_N into a Boolean algebra \mathfrak{B}_{N-1} of order 2^{N-1} on which Σ is, therefore, mapped. On repeating this process $N - 2$ times, we obtain a mapping upon the Boolean algebra \mathfrak{B}_2 . On retracing our steps from \mathfrak{B}_2 to \mathfrak{B}_3 to \mathfrak{B}_4 and so on to \mathfrak{B}_N , we see that we have a multiple isomorphism between \mathfrak{B}_N and \mathfrak{B}_2 which preserves the assertion values of all the postulates for Σ . Hence, the mapping on \mathfrak{B}_N is not essentially distinct from one of the three possible mappings on \mathfrak{B}_2 .

INSTITUTE FOR ADVANCED STUDY.

† The question of whether representations of Σ as a truth-value system of infinite order exist is left open.

‡ The reader may find it helpful to glance back at Table V. In the mapping over \mathfrak{B}_N , 1, 2, 3, 6, will be replaced by the 2^N elements of \mathfrak{B}_N . However, the elements on the extreme right of Table V which are all designated values of \mathfrak{B}_N , will remain designated values after equating α_1 and α_2 .

ON THE PROGRESSIONS ASSOCIATED WITH A TERNARY QUADRATIC FORM.

By E. H. HADLOCK.

1. *Introduction.* Denote the primitive ternary quadratic form $ax^2 + by^2 + cz^2 + 2xyz + 2syz + 2txy$ by f , its reciprocal by F , its Hessian or determinant by H , ($H \neq 0$), and the greatest common divisor of the cofactors of a, b, c etc., in H by Ω . Then Δ is defined by $H = \Omega^2 \Delta$.

B. W. Jones* has shown that with every ternary quadratic form f of Hessian H there is associated a set of arithmetic progressions:

$$(1) \quad 2^r(8n + a'_j), \quad p_i^{r_i}(p_i n + a_{ij}) \quad (n = 0, \pm 1, \pm 2, \dots)$$

such that no integer falling in any one of them is represented by f , and for every integer a not falling in any of them it is true that $f \equiv a \pmod{N}$, for N arbitrary,† is solvable, where p_i are odd prime factors of H , a_{ij} are some or all the members of a complete residue system mod p_i , r and r_i range over some or all of the positive integers and zero, and a'_j are some, none or all of 1, 3, 5, 7.

But in this paper we will speak of $2^r(8n + a'_j)$ as a set of progressions associated with f where a'_j is one of 1, 3, 5, or 7. Similarly, $p_i^{r_i}(p_i n + a_{ij})$ will be a set for each p_i of H .

In Art. I it is shown that Ω, Δ together with the order and the generic characters as defined by H. J. S. Smith‡ determine the progressions associated with a given form; and conversely, that Ω, Δ and the progressions associated with a given form determine the generic characters. In fact, it is important to notice that Ω and Δ restrict the choice of the progressions (1) as is seen on pp. 103-109. We shall speak of the progressions (1) associated with a given form as progressions corresponding to the generic characters and the invariants Ω and Δ of the form, or simply *corresponding progressions*. The corresponding progressions are given on pp. 103-109.

Smith§ has shown that there exists a properly primitive form f having

* B. W. Jones, "A new definition of genus for ternary quadratic forms," *Transactions of the American Mathematical Society*, vol. (1931), No. 1, pp. 92-110. This article will be referred to as Art. I.

† This condition implies that a is represented by some form of the same genus as f . (See B. W. Jones, "Regularity of a genus of positive ternary quadratic forms," *Transactions of the American Mathematical Society*, vol. 33 (1931), No. 1, pp. 111-124.)

‡ H. J. S. Smith, *Collected Mathematical Papers*, vol. 1, pp. 457-459; L. E. Dickson, "Studies in the Theory of Numbers," pp. 51, 52.

§ H. J. S. Smith, *loc. cit.*, p. 470.

a given Ω and Δ , a given set of values for the generic characters and whose reciprocal F is also properly primitive if and only if

$$(2) \quad \psi g^{ef} G^{eF} (f | \Omega_1) (F | \Delta_1) = (-1)^{e_2},$$

where

$$(2a) \quad \begin{aligned} g &= 1 \quad \text{if } \Omega = \Omega_1 \Omega_2^2, & g &= -1 \quad \text{if } \Omega = 2\Omega_1 \Omega_2^2, \\ G &= 1 \quad \text{if } \Delta = \Delta_1 \Delta_2^2, & G &= -1 \quad \text{if } \Delta = 2\Delta_1 \Delta_2^2, \\ e_2 &= (\Omega_1 + 1)(\Delta_1 + 1)/4, & e_F &= (P^2 - 1)/8, \end{aligned}$$

and Ω_2^2 and Δ_2^2 are the largest squares dividing Ω and Δ respectively. $\Omega = \Omega_1 \Omega_2^2$ or $2\Omega_1 \Omega_2^2$ according as Ω/Ω_2^2 is odd or even; similarly for Δ . Hence Ω_1 and Δ_1 are always odd and not divisible by any square. If f is improperly primitive then instead of (2) the condition is

$$(3) \quad (-G)^{eF} (2f | \Omega_1) (F | \Delta_1) = (-1)^{e_2}.$$

If F is improperly primitive, the condition is

$$(4) \quad (-g)^{ef} (f | \Omega_1) (2F | \Delta_1) = (-1)^{e_2}.$$

The purpose of this paper is to find conditions on the progressions associated with a form which are equivalent to Smith's character conditions (2)-(4). This leads to the fact that the number of sets of corresponding progressions of a certain kind is odd or even according as f is positive or indefinite. (See Theorem II of this paper). It is also found that with every positive form there are associated infinitely many progressions of numbers not represented by f .

2. From (2a) we notice that the odd primes which occur to even powers in both Ω and Δ do not affect the value of $(f | \Omega_1) (F | \Delta_1)$ in (2)-(4). Then, suppose we have given Ω , Δ and only the following sets of corresponding progressions I-X involving the distinct odd prime factors p_1, p_2, \dots, p_n which occur to odd powers in at least one of Ω and Δ . From Art. I, p. 103, it is seen that if we omit the progressions p_a^{2s+1} then I-X include all combinations of sets of corresponding progressions in p_1, p_2, \dots, p_n .

- | | |
|-------------------------------------------------|------------------------------------------------------------------------------------------|
| I. $p_{1j}^{2k}(p_{1j}n + \alpha_{1j}),$ | II. $p_{2j}^{2k}(p_{2j}n + \alpha_{2j}), p_{2j}^{2s_{2j}+1}(p_{2j}n + \beta_{2j}),$ |
| III. $p_{3j}^{2r_{3j}}(p_{3j}n + \alpha_{3j}),$ | IV. $p_{4j}^{2r_{4j}}(p_{4j}n + \alpha_{4j}), p_{4j}^{2s_{4j}+1}(p_{4j}n + \beta_{4j}),$ |
| V. $p_{5j}^{2k+1}(p_{5j}n + \beta_{5j}),$ | VI. $p_{6j}^{2r_{6j}}(p_{6j}n + \alpha_{6j}), p_{6j}^{2s_{6j}+1}(p_{6j}n + \beta_{6j}),$ |
| $p_{5j}^{2r_{5j}}(p_{5j}n + \alpha_{5j}),$ | |
| VII. $p_{7j}^{2k+1}(p_{7j}n + \beta_{7j}),$ | VIII. $p_{8j}^{2k+1}(p_{8j}n + \beta_{8j}), p_{8j}^{2r_{8j}}(p_{8j}n + \alpha_{8j}),$ |
| IX. $p_{9j}^{2r_{9j}}(p_{9j}n + \alpha_{9j}),$ | X. none for $p_{10j},$ |

where ($j = 1, 2, \dots, N_i$), ($i = 1, 2, \dots, 10$) and the p_{ij} 's are p_1, p_2, \dots, p_n renamed. Each α_{ij} and β_{ij} represents all the quadratic residues or all the quadratic non-residues of p_{ij} ; the ranges r_{ij} and s_{ij} are finite and that of k is infinite. In Art. I, p. 103, $\alpha_{-1}, \alpha_1, s, s_1, r, k, \Omega', \Delta'$ correspond to

$$(5) \quad \alpha_{ij}, \beta_{ij}, s'_{ij}, s_{ij}, r_{ij}, k, \Omega'_{ij}, \Delta'_{ij},$$

respectively. Ω'_{ij} and Δ'_{ij} are defined by

$$(6) \quad (\Omega/p_{ij}^{t_{ij}}) = \Omega'_{ij} \not\equiv 0 \pmod{p_{ij}}, \quad (\Delta/p_{ij}^{t'_{ij}}) = \Delta'_{ij} \not\equiv 0 \pmod{p_{ij}}.$$

In I-IV, p_{ij} occurs to odd and even powers in Ω and Δ respectively; in V and VI, p_{ij} occurs to odd powers in both Ω and Δ ; and in VII-X, p_{ij} occurs to an odd power in Δ and to an even power in Ω . p_{ij} of I and III is not a factor of Δ and p_{ij} of VII and X is not a factor of Ω .

Define

$$(7) \quad \Omega_{ii} = \prod_{j=1}^{N_i} p_{ij} \quad (i = 1, 2, \dots, 6), \quad \Delta_{ii} = \prod_{j=1}^{N_i} p_{ij} \quad (i = 5, \dots, 10),$$

$$(8) \quad A = \prod_{i=1}^4 \Omega_{ii}, \quad B = \prod_{i=7}^{10} \Delta_{ii},$$

$$(9) \quad C = \Omega_{55}\Omega_{66} = \Delta_{55}\Delta_{66},$$

$$(10) \quad J(u, v, w) = (-1 | uvw)(u | v)(v | u)(uv | w)(w | uv),$$

$$(11) \quad e_3 = N_1 + N_2 + N_5 + N_7 + N_8.$$

If $N_i = 0$, define $\Omega_{ii} = 1$, ($i = 1, 2, \dots, 6$) and $\Delta_{ii} = 1$, ($i = 5, \dots, 10$).

From (2a), (7), (8), and (9), we have

$$(12) \quad |\Omega_1| = \prod_{i=1}^6 \Omega_{ii} = AC, \quad |\Delta_1| = \prod_{i=5}^{10} \Delta_{ii} = BC.$$

Case 1. Ω and Δ are each positive.

Hence $\Omega_1 = AC$ and $\Delta_1 = BC$. From Lemma 12 of Art. I we notice that

$$(13) \quad (f | p) = (\alpha | p)$$

for each p of Ω and from the corollary of Lemma 4 we notice also that

$$(14) \quad (\alpha | p) = -(\alpha_{-1} | p).$$

With the aid of (13), (14), (5), and (7) we obtain

$$(15) \quad (f | \Omega_{ii}) = (-1)^{N_i} \prod_{j=1}^{N_i} (\alpha_{ij} | p_{ij}), \quad (i = 1, 2, \dots, 6).$$

From Art. I, p. 103 we have for each p in I-II, III-IV, VII-VIII, IX-X, V and VI respectively, the conditions

$$(16) \quad (-\alpha\Delta' | p) = -1, \quad (-\alpha\Delta' | p) = 1,$$

$$(17) \quad (F | p) = -(-\Omega' | p), \quad (F | p) = (-\Omega' | p),$$

$$(18) \quad (F | p) = -(-\alpha\Delta'\Omega' | p), \quad (F | p) = (-\alpha\Delta'\Omega' | p).$$

From (15) we have with the aid of (16), (6), and (2a)

$$(19) \quad (f | \Omega_{ii}) = \gamma^{N_i} G^{\epsilon_{\Delta_{ii}}} (-\Delta_1 | \Omega_{ii}), \quad (i = 1, 2, 3, 4),$$

where $\gamma = -1$ if $(i = 1, 2)$, $\gamma = 1$ if $(i = 3, 4)$. From (17) we have

$$(20) \quad (F | \Delta_{ii}) = \gamma^{N_i} g^{\epsilon_{\Delta_{ii}}} (-\Omega_1 | \Delta_{ii}), \quad (i = 7, \dots, 10),$$

where $\gamma = -1$ if $(i = 7, 8)$, $\gamma = 1$ if $(i = 9, 10)$. From (8), (9), and (12) we have

$$(21) \quad \begin{aligned} (\Delta'_{5j} | p_{5j}) &= G^{\epsilon_{p_{5j}}} (B\Delta_{66} | p_{5j}) (p_{51}, p_{52}, \dots, p_{5,j-1}, p_{5,j+1}, \dots, p_{5N_5} | p_{5j}), \\ (-\Omega'_{5j} | p_{5j}) &= g^{\epsilon_{p_{5j}}} (-A\Delta_{66} | p_{5j}) (p_{51}, p_{52}, \dots, p_{5,j-1}, p_{5,j+1}, \dots, p_{5N_5} | p_{5j}). \end{aligned}$$

Then from (18) and (21) we obtain

$$(22) \quad (F | \Delta_{55}) = (gG)^{\epsilon_{\Delta_{55}}} (-AB | \Delta_{55}) \prod_{j=1}^{N_5} (\alpha_{5j} | p_{5j}).$$

Similarly we obtain

$$(23) \quad (F | \Delta_{66}) = (-1)^{N_6} (gG)^{\epsilon_{\Delta_{66}}} (-AB | \Delta_{66}) \prod_{j=1}^{N_6} (\alpha_{6j} | p_{6j}).$$

From (15) and (19) we have

$$(24) \quad (f | \Omega_1) = (-1)^{e_4} G^{\epsilon_A} (-1 | A) (\Delta_1 | A) \prod_{j=1}^{N_5} (\alpha_{5j} | p_{5j}) \prod_{j=1}^{N_6} (\alpha_{6j} | p_{6j}),$$

where $e_4 = N_1 + N_2 + N_5 + N_6$. From (20), (22) and (23) we obtain

$$(25) \quad \begin{aligned} (F | \Delta_1) &= (-1)^{\epsilon_5} g^{\epsilon_{\Delta_1}} G^{\epsilon_C} (-1 | BC) (\Omega_1 | B) (AB | C) \prod_{j=1}^{N_5} (\alpha_{5j} | p_{5j}) \prod_{j=1}^{N_6} (\alpha_{6j} | p_{6j}) \end{aligned}$$

where $\epsilon_5 = N_6 + N_7 + N_8$. Then from (24), (25), (10) and (11) we have

$$(26) \quad (f | \Omega_1) (F | \Delta_1) = J(A, B, C) (-1)^{e_2} g^{\epsilon_{\Delta_1}} G^{\epsilon_{\Omega_1}}.$$

For the two cases, $A \equiv B \pmod{4}$, and $A \equiv 3B \pmod{4}$

$$(27) \quad J(A, B, C) (-1)^{e_2} = -1.$$

Case 2. Ω and Δ have opposite signs.

If $\Omega < 0$, then from (2a) we notice that $\Omega_1 < 0$. Then in (12) we take $\Omega_1 = (-A)C$. Instead of $J(A, B, C)$ in (26) we have $J(-A, B, C)$. But

$$(28) \quad J(-A, B, C)(-1)^{e_2} = 1.$$

If $\Omega > 0$ we also have $J(A, -B, C)(-1)^{e_2} = 1$.

Replace $(f | \Omega_1)(F | \Delta_1)$ in (2) by its value in (26) and use (27) and (28). We obtain

$$(29) \quad \psi g^{e_{\Delta_1} f} G^{e_{\Omega_1} F} = R(-1)^{e_3}$$

where $R = -1$ if Ω and Δ are each positive, and $R = 1$ if Ω and Δ have opposite signs. From the progressions I-X we notice that e_3 is the number of sets of corresponding progressions of the type $p_{ij}^{k_1}(p_{ij}n + C_{ij})$ where $C_{ij} = \alpha_{ij}$ or β_{ij} and $k_1 = 1, 3, 5, \dots$ or $k_1 = 0, 2, 4, \dots$; that is, the range of the exponent k_1 of p_{ij} is infinite. The condition (29) is equivalent to (2).

When we apply to (29) each of the cases A-F* in Art. I, pp. 104-108, we find that there exists a positive form or an indefinite form if and only if the number of sets of corresponding progressions

$$(30) \quad p_{ij}^{k_1}(p_{ij}n + C_{ij}), \quad 2^{k_2}(8n + a'_j)$$

is respectively odd or even, where k_i ($i = 1, 2$) ranges over all the evens or over all the odds. In particular we find from A that an indefinite form may have no progressions at all associated with it.

As an illustration of the application of A-E to (29) we take in E ($\Omega'' = 8, \Delta'' = 2$) the progressions $4n + \Delta', 4n + 2, 8n + 5\alpha\Delta', 4^k(16n + 14\Delta')$ which are found in the last row, with $\alpha \equiv 3 \pmod{4}$, and $\Omega_1 F \equiv \Omega' F \equiv 3\alpha, 5\alpha \pmod{8}$. From (29) we have $(2 | \Omega_1 F)(2 | \Delta_1 f)\psi = R(-1)^{e_3}$. If $\alpha \equiv 3 \pmod{8}$, then $f \equiv 3\Delta' \pmod{8}$; hence, $(2 | \Delta_1 f) = -1$, and $\psi = 1$. Also $(2 | \Omega_1 F) = 1$ since $\Omega_1 F \equiv 1, 7 \pmod{8}$. Then $R(-1)^{e_3} = -1$, and e_3 is even or odd according as $R = -1$ or 1 . Similarly, if $\alpha \equiv 7 \pmod{8}$.

For f in (3) $\Omega \equiv 1 \not\equiv \Delta \pmod{2}$. Now if we replace $(f | \Omega_1)(F | \Delta_1)$ in (3) by its value given in (26) we find that

$$(31) \quad \begin{aligned} R(-1)^{e_3} &= 1 & \text{if } G &= -1, \\ R(-1)^{e_3} &= (2 | \Omega_1 F) & \text{if } G &= 1. \end{aligned}$$

Then, from Art. I, p. 109, G, we find again that there exists a positive form or an indefinite form if and only if the number of sets of corresponding progressions (30) is respectively odd or even.

For f in (4) $\Delta \equiv 1 \not\equiv \Omega \pmod{2}$. Now if we replace $(f | \Omega_1)(F | \Delta_1)$ in (4) by its value given in (26) we find that

* $2^{k_2}(8n + a'_j)$ may be obtained directly from the congruences in III of Art. I, p. 104.

$$(32) \quad \begin{aligned} R(-1)^{e_s} &= 1 \quad \text{if } g = -1, \\ R(-1)^{e_s} &= (2 \mid \Delta_1 f) \quad \text{if } g = 1. \end{aligned}$$

If we apply the principle of reciprocity to (32) we again obtain (31). If now we display the progressions $4^k(8n + a'_j)$ of F in Art. I, p. 108, G, and apply (31) to them we find that we can have a form F and hence f having the corresponding progressions (30) if and only if the number of their sets is odd or even according as F is positive or indefinite.

Since (29), (31), and (32) are equivalent to (2), (3), and (4) respectively, we have

THEOREM I. *Each of Smith's character conditions (2)-(4) is equivalent to the fact that the number of sets of corresponding progressions* of the type $p_{ij}^{k_1}(p_{ij}n + C_{ij})$, $2^{k_2}(8n + a'_j)$ in the progressions (1) is odd or even according as Ω and Δ have the same or opposite signs, and where $C_{ij} = \alpha_{ij}$ or β_{ij} and k_i ($i = 1, 2$) ranges over all the evens or over all the odds.*

THEOREM II. *There is an odd (even) number of sets of progressions of the type $p_{ij}^{k_1}(p_{ij}n + C_{ij})$, $2^{k_2}(8n + a'_j)$ in the progressions (1) associated with a positive (indefinite) form where $C_{ij} = \alpha_{ij}$ or β_{ij} and k_i ($i = 1, 2$) ranges over all the evens or over all the odds. Conversely; if Ω and Δ are given and if there is given an odd (even) number of sets of corresponding progressions of the type $p_{ij}^{k_1}(p_{ij}n + C_{ij})$, $2^{k_2}(8n + a'_j)$ in the progressions (1), then there exists a positive (indefinite) form associated with the given progressions.*

COROLLARY. *With every positive ternary quadratic form there are associated infinitely many progressions (1).*

By inspection of p. 103 of Art. I we have the following properties: $P_1 - P_5$ where $L_{ij} = [(t_{ij} - 2)/2]$, $M_{ij} = [L_{ij} + t'_{ij}/2]$ and t_{ij} , t'_{ij} are defined in (6). $[-1/2] = -1$.

P_1 . If $t_{ij} \geq 2$ is even then $r_{ij} = 0, 1, 2, \dots, L_{ij}$, if t_{ij} is odd then $r_{ij} = 0, 1, 2, \dots, M_{ij} + 1$.

P_2 . If $t_{ij} \geq 2$ then $s'_{ij} = 0, 1, 2, \dots, L_{ij}$ unless $p_{ij}^{2k+1}(p_{ij}n + \beta_{ij})$ occurs when t_{ij} is even. If t_{ij} is even and $t_{ij} + t'_{ij} \geq 2$ then $s'_{ij} = 0, 1, 2, \dots, M_{ij}$.

P_3 . If $t_{ij} \geq 1$ is odd then $s_{ij} = 0, 1, 2, \dots, M_{ij}$ or $M_{ij} + 1$ according as t'_{ij} is even or odd.

P_4 . If t'_{ij} is odd then $(\beta_{ij} \mid p_{ij}) = (-\Delta'_{ij} \mid p_{ij})$ or $-(-\Delta'_{ij} \mid p_{ij})$ according as $p_{ij}^{2k+1}(p_{ij}n + \beta_{ij})$ or $p_{ij}^{2s_{ij}+1}(p_{ij}n + \beta_{ij})$ occurs; otherwise the value of $(\beta_{ij} \mid p_{ij})$ can be chosen arbitrarily.

* When F is improperly primitive the progressions (30) of f can readily be found.

P_5 . If t_{ij} is odd and t'_{ij} is even, then $(\alpha_{ij} | p_{ij}) = (-\Delta_{ij} | p_{ij})$ or $-(-\Delta_{ij} | p_{ij})$ according as $p_{ij}^{2k}(p_{ij}n + \alpha_{ij})$ on $p_{ij}^{2r_{ij}}(p_{ij}n + \alpha_{ij})$ occurs; otherwise the value of $(\alpha_{ij} | p_{ij})$ can be chosen arbitrarily.

3. *Example.* We now give an example of the converse part of Theorem II. Suppose we have given $\Omega = 6$, $\Delta = 105$. Since 3 occurs to an odd power in both Ω and Δ , we may choose either V or VI. Choose $3^{2k+1}(3n + \beta_{51})$, $3^{r_{51}}(3n + \alpha_{51})$. From P_4 , $(\beta_{51} | 3) = (-35 | 3) = 1$. From P_1 we notice that $r_{51} = 0$ and from P_5 we may choose $\alpha_{51} \equiv 2 \pmod{3}$. Since 5 is not a factor of Ω and occurs to an odd power in Δ we may choose either VII or X. Choose $5^{2k+1}(5n + \beta_{71})$. $(\beta_{71} | 5) = (-21 | 5) = 1$. Let there be no progressions involving 7. There are no progressions $p^{2s+1}a$ for $p = 3, 5$, and 7. The number of sets of progressions $3^{2k+1}(3n + 1)$ and $5^{2k+1}(5n \pm 1)$ is even. Hence in Art. I, p. 104 ($\Omega'' = 2$, $\Delta'' = 1$) we must take $4^k(8n + 7\Delta')$. $\Delta' = \Delta \equiv 1 \pmod{8}$.

It remains to find a form f with $\Omega = 6$, $\Delta = 105$, and having the progressions $3^{2k+1}(3n + 1)$, $3n + 2$, $5^{2k+1}(5n \pm 1)$ and $4^k(8n + 7)$ associated with it.

From $f = ax^2 + by^2 + cz^2 + 2xyz + 2xz$ we obtain $abf = bx_1^2 + ay_1^2 + Hz^2$ where $x_1 = ax + z$, $y_1 = by + rz$ and $H = a\Omega A - b$. Hence $b \equiv 0 \pmod{\Omega}$. Take $b = 6b'$ with b' prime to 6. $g = 2ab'f = 2b'x_1^2 + 3ay_1^2 + 1260z^2$ where $y_1 = 3y_2$. In order for f to have the progression $3n + 2$ associated with it, we take $a \equiv 1 \pmod{3}$. $g \equiv 0 \pmod{3}$ implies that $x_1 = 3x_2$. Then $g/3 = 6b'x_2^2 + ay_2^2 + 420z^2$. We take $b' \equiv 1 \pmod{3}$. Then f will have the progressions $3^{2k+1}(3n + 1)$ associated with it as is seen from the corollary of Lemma 5 of Art. I. Similarly, we take $(ab' | 5) = -1$ and $(ab' | 7) = 1$. Let $a = 1$. $b'f = b'x_1^2 + 6y_2^2 + 630z^2$ where $y_2 = 2y_3$. Let $b' \equiv 3 \pmod{8}$. Then from the corollary of Lemma 11 of Art. I, f will have the progressions $4^k(8n + 7)$ associated with it. Take $b' = 67$. Then $b = 402$. From $H = a(bc - r^2) - b$ we have $c = 133$, and $r = 222$. Hence $f = x^2 + 402y^2 + 133z^2 + 444yz + 2xz$. From $x_1 = x + z$, x will be an integer for any integers assigned to x_1 and z . $y_1 = by + rz = 6y_3$ where $y_3 = 67y + 37z$. $b'f \equiv 0 \pmod{67}$ implies that $y_3 \equiv \pm 37z \pmod{67}$. Hence the sign of z can be so chosen that y is an integer. Apply to f the transformation $x = x' + 2y' + z'$, $y = y' + z'$ and $z = -2y' - z'$ of determinant 1. We find that f is equivalent to the reduced form $x'^2 + 42y'^2 + 90z'^2$.

4. *Table.* In the following table * abbreviate the form f by enclosing the

* Eisenstein has given a table of genera for forms with odd Hessian from 1 to 25, "neue Theoreme der höheren Arithmetik," *Journal für Mathematik*, vol. 35 (1847), p. 136.

coefficients of the square terms and half the coefficients of the product terms in parentheses: (a, b, c, r, s, t) . Let P denote the progressions (1) associated with a form f . Let $s_1 = (2 | f)\psi$ and $s_2 = (2 | f)(2 | F)\psi$. An asterisk prefixed to a form indicates that f has an improperly primitive reciprocal F . If f is improperly primitive, then f always has the progression $2n + 1$ associated with it. The progression $2n + 1$ is not written in the table.

TABLE OF GENERIC CHARACTERS AND PROGRESSIONS OF REDUCED POSITIVE
TERNARY QUADRATIC FORMS FOR TYPICAL VALUES OF H FROM 1 TO 25.

<i>f</i> Properly Primitive.				
H odd.				
H	$(F p)$ or $(f p)$	ψ	P	Forms
1.		—1	$4^k(8n + 7)$	$(1, 1, 1, 0, 0, 0)$
3	1	1	$3^{2k+1}(3n + 2)$	$(1, 1, 3, 0, 0, 0)$
3	—1	—1	$4^k(8n + 5)$	$(1, 2, 2, -1, 0, 0)$
5	1	—1	$4^k(8n + 3)$	$(1, 1, 5, 0, 0, 0)$
5	—1	1	$5^{2k+1}(5n \pm 1)$	$(1, 2, 3, -1, 0, 0)$
9	1	—1	$4^k(8n + 7)$	$(1, 1, 9, 0, 0, 0)$
			$3(3n \pm 1)$	
9	—1	—1	$4^k(8n + 7)$	$(1, 2, 5, -1, 0, 0)$
9	$(f 3) = 1$	1	$3^{2k}(3n + 2)$	$(1, 3, 3, 0, 0, 0)$
9	$(f 3) = -1$	—1	$4^k(8n + 7)$	$(2, 2, 3, 0, 0, -1)$
			$3n + 1$	
15	1	1	$3^{2k+1}(3n + 1)$	$(1, 1, 15, 0, 0, 0)$
	1			$(1, 4, 4, -1, 0, 0)$
15	—1	1	$5^{2k+1}(5n \pm 2)$	$(1, 2, 8, -1, 0, 0)$
	—1			$(1, 3, 5, 0, 0, 0)$
15	$(F 3) = 1$	—1	$4^k(8n + 1)$	$(2, 2, 5, 0, 0, -1)$
	$(F 5) = -1$		$5^{2k+1}(5n \pm 2)$	
			$3^{2k+1}(3n + 1)$	
15	$(F 3) = -1$	—1	$4^k(8n + 1)$	$(2, 3, 3, 0, 0, -1)$
	$(F 5) = 1$			
$H \equiv 2 \pmod{4}$				
H	$(F p)$ or $(f p)$	$(2 F)\psi$	P	Forms
2		—1	$4^k(16n + 14)$	$(1, 1, 2, 0, 0, 0)$
6	1	1	$3^{2k+1}(3n + 1)$	$(1, 1, 6, 0, 0, 0)$
6	—1	—1	$4^k(16n + 10)$	$(1, 2, 3, 0, 0, 0)$
18	1	—1	$4^k(16n + 14)$	$(1, 1, 18, 0, 0, 0)$
			$3(3n \pm 1)$	$(2, 2, 5, 0, -1, 0)$
18	—1	—1	$4^k(16n + 14)$	$(1, 2, 9, 0, 0, 0)$
				$(2, 3, 4, -1, 0, -1)$
18	$(f 3) = 1$	1	$4^k(16n + 14)$	$(1, 3, 6, 0, 0, 0)$
			$3n + 2$	
18	$(f 3) = -1$	—1	$3^{2k}(3n + 1)$	$(2, 3, 3, 0, 0, 0)$

$H \equiv 0 \pmod{4}$					
$f \equiv F \pmod{8}$					
or					
H	$(F p)$	s_1	s_2	P	Forms
4		1, 5	1, 5	$4^k(8n+7)$ $4n+3$	$(1, 1, 4, 0, 0, 0)$
4		$s_1 = -1$		$4^k(8n+7)$	$(1, 2, 2, 0, 0, 0)$
8		1, 5	1	$4^k(16n+14)$ $8n+6$ $4n+3$	$(1, 1, 8, 0, 0, 0)$
8		$s_2 = -1$		$4^k(16n+14)$	$(1, 2, 4, 0, 0, 0)$
8			3	$4^k(16n+14)$ $4n+2$	$(1, 3, 3, -1, 0, 0)$
8		3, 7	5	$4^k(16n+14)$ $8n+6$ $4n+1$	$(2, 2, 3, -1, -1, 0)$
12	1	1, 5	1, 5	$3^{2k+1}(3n+2)$ $4n+3$	$(1, 1, 12, 0, 0, 0)$
12	1	$s_1 = 1$		$4^k(8n+5)$	$(1, 2, 6, 0, 0, 0)$
12	1		3, 7	$3^{2k+1}(3n+2)$ $4n+2$	$(1, 3, 4, 0, 0, 0)$
12	-1	1, 5		$3^{2k+1}(3n+2)$ $4n+3$ $4n+2$	$*(1, 4, 4, -2, 0, 0)$
12	-1	$s_1 = -1$		$3^{2k+1}(3n+2)$ $8n+1$	$(2, 2, 3, 0, 0, 0)$
12	-1	3, 7	1, 5	$4^k(8n+5)$ $8n+1$	$(2, 3, 3, 1, 1, 1)$

f Improperly Primitive.

$(F p)$					
or					
H	$(f p)$	$F \equiv$ $\pmod{8}$	P	Forms	
4		3	$4^k(8n+7)$	$(2, 2, 2, 1, 1, 1)$	
6	1	3, 7	$3^{2k+1}(3n+1)$	$(2, 2, 2, 0, 0, -1)$	
12	1	7	$3^{2k+1}(3n+2)$	$(2, 2, 4, -1, -1, 0)$	
12	-1	3	$4^k(8n+5)$	$(2, 2, 4, 0, 0, -1)$	
18	$(f 3) = -1$	1, 5	$3^{2k}(3n+1)$	$(2, 2, 6, 0, 0, -1)$	

A DEFINITION OF GROUP BY MEANS OF THREE POSTULATES.

By RAYMOND GARVER.

Given a set of elements $G(a, b, c, \dots)$ and a rule of combination, which may be called multiplication, by which any two elements of G , whether they be the same or different, taken in a specified order, determine a unique result, or product, which may or may not be an element of G . This system is called a group if it satisfies certain postulates; the sets of postulates to which we shall have occasion to refer in the present paper are chosen from the following list:

- I. (Closure). If a and b are elements of G , the product ab is an element of G .
- II. (Associativity). If $a, b, c, ab, bc, (ab)c, a(bc)$ are elements of G , then $(ab)c = a(bc)$.
- III. (Strengthened Associativity). If $a, b, c, ab, bc, (ab)c$ are elements of G , then $(ab)c = a(bc)$.
- IV. If a and b are elements of G , there exists an element x of G such that $ax = b$.
- V. If a and b are elements of G , there exists an element y of G such that $ya = b$.
- VI. (Existence of right-hand identity element). There exists an element e of G such that, for every element a of G , $ae = a$.
- VII. (Existence of right-hand inverse element). If such elements e occur, then for a particular e and for every element a of G there exists an element a' of G such that $aa' = e$.

One important definition of group employs I, II, VI and VII. This formulation is due to Dickson (ref. 11), but it is based to a large extent on the work of Moore. These four postulates were proved by Dickson to be independent. It is worth while pointing out that postulate VII must be stated carefully; van der Waerden's form of this postulate (ref. 12) is ambiguous, as Clifford has shown (ref. 13).†

The reader is familiar with the fact that this postulate system is often met in a slightly different form, with VI and VII replaced by stronger statements which postulate identity and inverse elements, not merely *right-hand* identity and inverse elements. Dickson's work shows, of course, that it is not necessary to postulate these stronger statements.

* The symbols a, b, \dots , as used in the postulates, need not represent *distinct* elements of G .

† Instead of VI and VII, van der Waerden postulates left-hand, instead of right-hand, identity and inverse elements. This is essentially the same type of definition.

The other commonly used definition of group makes use of I, II, IV and V. This set is a simplification of that used by Weber (ref. 1). He first defined a finite group by means of I, II and two other postulates whose exact form does not interest us here, and then deduced IV and V, with the uniqueness of the x and y there appearing, as theorems for finite groups. Finding that IV and V could not be so deduced for infinite groups, he added them, plus uniqueness postulates for x and y , to his set of postulates to define an infinite group. While this was a perfectly natural step to take, it led to a number of redundancies. Huntington, in 1902, showed (ref. 3) that Weber's other postulates could be deduced from I, II, IV, and V, but he did not actually emphasize having done so until 1905 (ref. 10). Moore, also in 1902, was the first to set up and study (ref. 5) the precise set I, II, IV, V.

Moore, however, left open the question as to whether I, II, IV and V form an independent set of postulates (ref. 5, page 489). I have recently been able to prove (ref. 14) that they are not independent; in fact, the closure property I can be deduced from II, IV, and V. This gives a simple definition of group by means of three postulates; further, no other postulate in the set I, II, IV and V can be deduced from the remaining three, as I have easily shown. That is, there is only one permissible three-postulate definition of group, if the three are to be chosen from I, II, IV, and V, and this set II, IV, and V is an independent set.

It should be mentioned that an earlier definition of group by three postulates was given by Huntington in 1902 (refs. 3 and 6). He employed III, IV, and V, his proof requiring the strengthened form of the associativity postulate. This definition may be thought of as the next to last step in the simplification of Weber's set of essentially 8 postulates to the set II, IV, V.

In this paper I propose to prove that a group may be defined by means of the three postulates II, IV and VI. While VI is, of course, not a weakened form of V in the sense that II is a weakened form of III, I think it will be generally agreed that VI is a "weaker" postulate than V. To justify this statement, assume that the multiplication table of the elements of G is given by means of a square array, whether finite or infinite in extent:

	a	b	c	\dots
a	p_{11}	p_{12}	p_{13}	\dots
b	p_{21}	p_{22}	p_{23}	\dots
c	p_{31}	p_{32}	p_{33}	\dots
\cdot	\cdot	\cdot	\cdot	\dots
\cdot	\cdot	\cdot	\cdot	\dots
\cdot	\cdot	\cdot	\cdot	\dots

The products p_{ij} may or may not be elements of G . Now we see that postulate V may be thought of as a restriction on *every* column of this square array; it requires *each* column to contain *every* element of G . On the other hand, postulate VI merely restricts one column of the array; there must exist an index i such that the column of products $p_{1i}, p_{2i}, p_{3i}, \dots$ is identical with the left border of the table a, b, c, \dots . There is, I think, then some justification for the belief that the definition of group by postulates II, IV and VI is the most satisfactory, from the logical standpoint, which has yet been given.

It may be pointed out that, in the light of VI, IV may be weakened slightly by the addition of the hypothesis $a \neq b$. This is hardly an important change. It may further be noted that IV and VI may be replaced by the composite postulate

VIII. If a and b are elements of G , there exists an element x of G such that $ax = b$; if $b = a$, there exists an element e of G such that, for any a in G , we may take $x = e$.

This does not in any real sense afford a reduction to two postulates, but it does emphasize an interesting relation between IV and VI.

To prove that a group may be defined by postulates II, IV, VI, we first deduce I and then V. The deduction of V is sufficient, since, as pointed out above, I have already obtained I as a consequence of II, IV, V; I am unable, however, to obtain V without obtaining I.

Assume, then, that a and b are elements of G . We wish to show that the product ab lies in G .

- (1) By VI, $\exists e$ in G such that $ae = a$.
- (2) By VI, $ee = e$.
- (3) By IV, $\exists c$ in G such that $ec = a$.
- (4) By IV, $\exists d$ in G such that $ed = c$.
- (5) By (2) and (4), $(ee)d = ed = c$.
- (6) By (4) and (3), $e(ed) = ec = a$.
- (7) By (5), (6) and II, $c = d$.
- (8) By (3) and (7), $ea = a$, for any a in G .
- (9) By IV, $\exists a'$ in G such that $aa' = e$.
- (10) By IV, $\exists a''$ in G such that $a'a'' = e$.
- (11) By (9) and (8), $(aa')a'' = ea'' = a''$.
- (12) By (10) and (1), $a(a'a'') = ae = a$.
- (13) By (11), (12) and II, $a'' = a$.
- (14) By (13) and (10), $a'a = e$.

- (15) By IV, $\exists f$ in G such that $a'f = b$.
 (16) By IV, $\exists g$ in G such that $ag = f$.
 (17) By (14) and (8), $(a'a)g = eg = g$.
 (18) By (16) and (15), $a'(ag) = a'f = b$.
 (19) By (17), (18) and II, $g = b$.
 (20) By (19) and (16), $ab = f$.

Since the product $ab = f$, an element of G , property I is established. Property V then follows at once, for, if we take $y = ba'$,

- (21) By II, (14) and VI, $ya = (ba')a = b(a'a) = be = b$.

We have thus exhibited an element y which satisfies V.

It is not without interest to point out that two important group properties follow easily from intermediate steps in the above proof, before closure has been deduced. Thus, at the end of step (8), we have proved that any right-hand identity element is also a left-hand identity element. Thus there exists an identity element e such that, for every element a of G , $ae = ea = a$. It then follows at once, by a familiar step, that there is a unique identity element. From (9) and (14) above we have, if a is in G , the existence of an a' in G such that $aa' = a'a = e$, in other words, the existence of an inverse element. It follows easily that, for a given a in G , the inverse a' is unique.

One question of some interest remains. If postulates II, IV and VI are sufficient to define a group, as we have showed, is the same true for the set of postulates II, V and VI? The answer is no; the simplest example of a system satisfying these postulates and yet not forming a group is given by the multiplication table

	e	a
e	e	e
a	a	a

The set of postulates II, V and VI is related to the concept of multiple group, as defined by Clifford (ref. 13). One of the two types of multiple group, the two types not being essentially different, satisfies II, V, with uniqueness of the element y there appearing, VI, and, in addition, I. But Clifford shows that a multiple group is not, in general, a group.

Finally, postulates II, IV and VI are independent and completely independent when the number of elements in G is greater than two. (When the number of elements is two, II is a consequence of IV and VI.) Examples to prove this can be written down easily.

BIBLIOGRAPHY.

1. Weber, *Lehrbuch der Algebra*, vol. 2 (1896), pp. 3-4.
2. Pierpont, *Annals of Mathematics*, (2), vol. 2 (1900), p. 47.
3. Huntington, *Bulletin of the American Mathematical Society*, vol. 8 (1902), pp. 296-300.
4. Huntington, *ibid.*, pp. 388-391.
5. Moore, *Transactions of the American Mathematical Society*, vol. 3 (1902), pp. 485-492.
6. Huntington, *Transactions of the American Mathematical Society*, vol. 4 (1903), p. 30.
7. Moore, *Transactions of the American Mathematical Society*, vol. 5 (1904), p. 549.
8. Huntington, *Transactions of the American Mathematical Society*, vol. 6 (1905), pp. 34-35.
9. Moore, *ibid.*, pp. 179-180.
10. Huntington, *ibid.*, pp. 181-197.
11. Dickson, *ibid.*, pp. 198-204.
12. van der Waerden, *Moderne Algebra*, vol. 1 (1930), p. 15.
13. Clifford, *Annals of Mathematics*, (2), vol. 34 (1933), pp. 865-871.
14. Garver, *Bulletin of the American Mathematical Society*, vol. 40 (1934), pp. 698-701.

UNIVERSITY OF CALIFORNIA AT LOS ANGELES.

THE SIMULTANEOUS REDUCTION OF TWO MATRICES TO TRIANGLE FORM.

By J. WILLIAMSON.

Introduction. A square matrix $A_0 = (a_{ij})$, $(i, j = 1, 2, \dots, n)$, whose elements a_{ij} are complex numbers is said to be a *triangle-matrix*, if $a_{ij} = 0$, when $i > j$, or, in other words, if each element to the left of the leading diagonal is zero. The elements a_{ii} , $(i = 1, 2, \dots, n)$, of the leading diagonal of a triangle-matrix A_0 are the latent roots or characteristic numbers of A_0 . Since the sum, the difference and the product of any two triangle-matrices are all triangle-matrices, if $f(A_0, B_0) = C_0$ is a matrix polynomial in the two matrices A_0 and B_0 , C_0 is a triangle-matrix. In particular, if

$$(1) \quad \begin{aligned} C_0 &= (c_{ij}), & (i, j &= 1, 2, \dots, n), \\ c_{ii} &= f(a_{ii}, b_{ii}), & (i &= 1, 2, \dots, n), \end{aligned}$$

so that the latent root c_{ii} of C_0 is the same function of the latent roots a_{ii} and b_{ii} , that C_0 is of A_0 and B_0 . Moreover, if A and B are similar to A_0 and B_0 respectively, so that there exists a non-singular matrix X satisfying the two equations

$$XA_0X^{-1} = A \text{ and } XB_0X^{-1} = B,$$

then

$$Xf(A_0, B_0)X^{-1} = f(A, B) = C,$$

and the latent roots of A, B, C are the latent roots of A_0, B_0, C_0 respectively. Consequently equation (1) is true when a_{ii}, b_{ii} and c_{ii} are the latent roots respectively of A, B and C .

Now, if $D_0 = \phi(A_0, B_0)$ is a second polynomial in the matrices A_0 and B_0 , and if, when x and y are indeterminates, $\phi(x, y) \equiv f(x, y)$, $C_0 - D_0$ is a triangle-matrix whose leading diagonal is zero. For, the element in the i -th place of the leading diagonal of this matrix is,

$$c_{ii} - d_{ii} = f(a_{ii}, b_{ii}) - \phi(a_{ii}, b_{ii}) = 0.$$

Consequently $(C_0 - D_0)^n = 0$; that is, the matrix $C_0 - D_0$ is nilpotent. Hence, if it is possible to reduce the two matrices A and B to triangle form by the same similarity transformation the matrix $f(A, B) - \phi(A, B)$ is nilpotent for every pair of polynomials f and ϕ , which satisfy the identity $f(x, y) \equiv \phi(x, y)$.

In what follows we shall be interested in the converse of this last statement. In particular we shall show that, if certain restrictions are placed on the matrix A , a sufficient condition that it be possible to reduce A and B to triangle form by the same unitary transformation is that a finite number of matrices, each of the form $h(A)(AB - BA)$, where $h(A)$ is a polynomial in A , be nilpotent.

We shall have occasion to write an n -rowed square matrix S as a matrix of matrices,

$$(2) \quad S = (S_{ij}), \quad (i, j = 1, 2, \dots, t),$$

where S_{ij} is a matrix of e_i rows and e_j columns. If T is a second n -rowed square matrix and

$$(3) \quad T = (T_{ij}), \quad (i, j = 1, 2, \dots, t),$$

where T_{ij} is a matrix of e_i rows and e_j columns, we shall say that S and T are *similarly partitioned* or that (3) is a partition of T similar to (2). If in (2) $S_{ij} = 0$, when $i \neq j$, we shall call S a *diagonal block* matrix and write

$$(4) \quad S = [S_1, S_2, \dots, S_t],$$

where $S_i = S_{ii}$, $i = 1, 2, \dots, t$.

We shall use E to denote the unit matrix and U to denote the *auxiliary unit* matrix, whose only non-zero elements lie in the diagonal above the leading one, each of which is unity.*

1. Let A be a square matrix of order n over the field of all complex numbers and let the elementary divisors of $A - \lambda E$ be

$$(\lambda - \lambda_1)^{e_1}, (\lambda - \lambda_2)^{e_2}, \dots, (\lambda - \lambda_t)^{e_t},$$

where $e_i \leq 1$ and $e_1 + e_2 + \dots + e_t = n$. The classical canonical form of A is the diagonal block matrix,†

$$(5) \quad A_n = [M_1, M_2, \dots, M_t].$$

In (5) M_i is a square matrix of order e_i ; in fact

$$(6) \quad M_i = \lambda_i E_i + U_i,$$

where E_i is the unit matrix of order e_i and U_i the auxiliary unit matrix of the same order. The matrix

$$(7) \quad h(A_n) = (A_n - \lambda_1 E)^{r_1} (A_n - \lambda_2 E)^{r_2} \dots (A_n - \lambda_t E)^{r_t}, \quad r_i \geq 0,$$

* Cf. Turnbull and Aitken, *Canonical Matrices*, p. 62.

† Dickson, *Modern Algebraic Theories*, p. 106.

is a polynomial in A_n and is a diagonal block matrix $[N_1, N_2, \dots, N_t]$, where

$$(8) \quad N_i = v_{ir_i} U_i^{r_i}, \quad (i = 1, 2, \dots, t),$$

and

$$(9) \quad \begin{aligned} v_{ir_i} &= 0, \quad r_i \geq e_i, \\ v_{ir_i} &= (\lambda_i - \lambda_1)^{r_1} (\lambda_i - \lambda_2)^{r_2} \dots (\lambda_i - \lambda_{i-1})^{r_{i-1}} (\lambda_i - \lambda_{i+1})^{r_{i+1}} \dots (\lambda_i - \lambda_t)^{r_t}, \quad r_i < e_i. \end{aligned}$$

In (7) it is understood that, if $r_i = 0$, the factor $(A_n - \lambda_i E)^{r_i}$ is replaced by the identity matrix. Let

$$(10) \quad B_n = (B_{ij}), \quad (i, j = 1, 2, \dots, t),$$

be a partition of the matrix B_n similar to that of A_n in (5). Then, if $A_n B_n - B_n A_n = C$ and $C = (C_{ij})$ is a partition of C , similar to that of B_n in (10),

$$\begin{aligned} C_{ij} &= M_i B_{ij} - B_{ij} M_j \\ &= (\lambda_i E_i + U_i) B_{ij} - B_{ij} (\lambda_j E_j + U_j) \end{aligned}$$

or

$$(11) \quad C_{ij} = (\lambda_i - \lambda_j) B_{ij} + U_i B_{ij} - B_{ij} U_j, \quad (i, j = 1, 2, \dots, t).$$

We shall find it convenient to use the notation $b(i, j; r, s)$ for the element in the r -th row and s -th column of the matrix B_{ij} and more generally $f(i, j; r, s)$ for the element in the r -th row and s -th column of the matrix F_{ij} , where $F = (F_{ij})$ is a partition of a matrix F similar to that of B_n in (10). With this notation equation (11) becomes

$$(12) \quad c(i, j; r, s) = (\lambda_i - \lambda_j) b(i, j; r, s) + b(i, j; r+1, s) - b(i, j; r, s-1),$$

$$(i, j = 1, 2, \dots, t; r = 1, 2, \dots, e_i; s = 1, 2, \dots, e_j),$$

with the understanding that $b(i, j; e_i + 1, s) = b(i, j; r, 0) = 0$.

We now make two hypotheses;

(a) *The matrix A is not derogatory*; that is the minimum equation satisfied by A is of degree n ;

(b) *For every polynomial $h(A_n)$ defined by (7), where $r_i = 0, 1, 2, \dots, e_i$ and $r_1 + r_2 + \dots + r_t \leq n - 2$, the matrix $h(A_n)(A_n B_n - B_n A_n)$ is nilpotent.*

As a consequence of hypothesis (a) we see that the latent root λ_i of A_n is distinct from the latent root λ_j , if $i \neq j$, and accordingly that each v_{ir_i} in (9) is different from zero, when r_i is less than e_i .

Now, if

$$h(A_n)(A_n B_n - B_n A_n) = h(A_n)C = F$$

and, if $F = (F_{ij})$ is a partition of F similar to that of B_n in (10),

$$(13) \quad F_{ij} = N_i C_{ij} = v_{ir_i} U_{ir_i} C_{ij}, \quad (i, j = 1, 2, \dots, t)$$

and

$$(14) \quad f(i, j; r, s) = v_{ir_i} c(i, j; r + r_i, s), \\ (i, j = 1, 2, \dots, t; r = 1, 2, \dots, e_i; s = 1, 2, \dots, e_j),$$

where $c(i, j; r + r_i, s) = 0$, if $r + r_i > e_i$.

We shall now prove

LEMMA I. *If i_1, i_2, \dots, i_p , is a subsequence of the sequence $1, 2, \dots, t$ and $p \geq 2$, then*

$$(15) \quad b(i_1, i_2; s_1, 1) b(i_2, i_3; s_2, 1) \cdots b(i_p, i_1; s_p, 1) = 0,$$

for all positive integers $s_j \leq e_j$.

To prove this lemma we first show that

$$(16) \quad c(i_1, i_2; s_1, 1) c(i_2, i_3; s_2, 1) \cdots c(i_p, i_1; s_p, 1) = 0,$$

for all values of $p = 1, 2, \dots, t$. We shall prove (16) by induction, assuming it true for $p = 1, 2, \dots, h - 1$ and to simplify our notation shall write m_j for e_{i_j} .

If (16) is not true when $p = h$, for some set of integers $q_j \leq m_j$ the product

$$(17) \quad g = c(i_1, i_2; q_1, 1) c(i_2, i_3; q_2, 1) \cdots c(i_h, i_1; q_h, 1)$$

is different from zero. Moreover, if α is a positive integer and $c(i_j, i_{j+1}, q_j + \alpha, 1)$ is different from zero, q_j may be replaced by $q_j + \alpha$ in g , and the resulting product will still be different from zero. Hence we may so choose the integers q_j in (17), that

$$(18) \quad c(i_j, i_{j+1}, q_j + \alpha, 1) = 0, \quad \alpha > 0,$$

$$(19) \quad c(i_j, i_{j+1}, q_j, 1) \neq 0, j = 1, 2, \dots, h; i_{h+1} = i_1.$$

If $k \not\equiv j + 1 \pmod{h}$ and $s_j \leq m_j$, one of the products

$$c(i_j, i_k; s_j, 1) c(i_k, i_{k+1}; q_k, 1) \cdots c(i_{j-1}, i_j; q_{j-1}, 1)$$

or

$$\bullet \quad c(i_j, i_k; s_j, 1) c(i_k, i_{k+1}; q_k, 1) \cdots c(i_h, i_1; q_h, 1) \cdots c(i_{j-1}, i_j; q_{j-1}, 1)$$

is zero by our induction assumption, since it is of type (16) with $p \leq h - 1$. Therefore by (19)

$$(20) \quad c(i_j, i_k; s_j, 1) = 0, \\ (j, k = 1, 2, \dots, h; \quad k \not\equiv j + 1 \pmod{h}, s_j = 1, 2, \dots, m_j).$$

Let $h(A_n)$ be the polynomial defined by (7) for which $r_k = e_k$, if k does not lie in the set i_1, i_2, \dots, i_h , and $r_k = q_j - 1$, if $k = i_j$. Then, if we write v_j for $v_{i_{q_j-1}}$, it follows from (9) and hypothesis (a) that v_1, v_2, \dots, v_h are all different from zero while all other v_{ik} in $h(A_n)$ are zero. Hence by (13), for this particular polynomial $h(A_n)$,

$$(21) \quad F_{ij} = 0, \quad j = 1, 2, \dots, t; \quad i \text{ not in the set } i_1, i_2, \dots, i_h,$$

and by (14) and (20),

$$(22) \quad f(i_j, i_k; s_j, 1) = v_j c(i_j, i_k; s_j + q_j - 1, 1) = 0, \quad k \not\equiv j + 1 \pmod{h},$$

while by (14), (18) and (19)

$$(23) \quad \begin{cases} f(i_j, i_{j+1}; 1, 1) = v_j c(i_j, i_{j+1}; q_j, 1) \neq 0 \\ f(i_j, i_{j+1}; s_j, 1) = v_j c(i_j, i_{j+1}; s_j + q_j - 1, 1) \\ \quad = 0, \quad j = 1, 2, \dots, h; \quad i_{h+1} = i_1, s_j \geq 2. \end{cases}$$

Since by hypothesis (b) the matrix F is nilpotent, so is the matrix $H = F^h$. If $H = (H_{ij})$ is a partition of H similar to that of F ,

$$(24) \quad H_{ij} = F_{i\alpha_2} F_{\alpha_2\alpha_3} \cdots F_{\alpha_h j},$$

where each α_i is summed from 1 to t . It follows from (21), that each α_i need only be summed over the set i_1, i_2, \dots, i_h and that H_{ij} is zero, if i does not lie in the set i_1, i_2, \dots, i_h . Consequently H is nilpotent, if, and only, if the matrix

$$(25) \quad Q = (H_{ij_{i_k}}), \quad (j, k = 1, 2, \dots, h),$$

is nilpotent. Moreover every matrix in the product $F_{i_1\alpha_2} F_{\alpha_2\alpha_3} \cdots F_{\alpha_h i_k}$, as a consequence of (22) and (23), is a matrix, whose first column is zero, except, perhaps, for the element in the first row. The same is therefore true of the product matrix and the element in the first row and column of this matrix is,

$$W = f(i_j, \alpha_2; 1, 1) f(\alpha_2, \alpha_3; 1, 1) \cdots f(\alpha_h, i_k; 1, 1).$$

But, by (22) and (23), W is different from zero if, and only if, $i_j = i_k$ and $\alpha_s = i_{j+s-1}$. Hence every element in the first column of $H_{ij_{i_k}}$ is zero, if $k \neq j$, and, consequently, every element in the first column of Q , defined by (25), except the element in the first row, is zero. The element in the first row and first column of Q is,

$$h(i_1, i_1; 1, 1) = v_1 v_2 \cdots v_h g$$

by (23) and (14). Since Q is nilpotent, $v_1 v_2 \cdots v_h g = 0$ and, as $v_1 v_2 \cdots v_h$ is not zero, g must be zero.

This contradiction shows that, if (16) is true when $p = h - 1$, it is also true when $p = h$. A repetition of the above argument with $h = 1$ and H replaced by F shows that (16) is true when $h = 1$, so that our proof by induction is complete and (16) is true for all values of $p \leq t$.

In proving (16) by induction from $h - 1$ to h we use certain polynomials $h(A_n)$. Of the exponents r_i in these polynomials $h(A_n)$ only $t - h$ have their maximum value e_i , so that, if $h \geq 2$, the sum $r_1 + r_2 + \cdots + r_t$ is at most $n - 2$. In the proof for $h = 1$, every r_i except one has its maximum value e_i ; but, since,

$$\begin{aligned} c(i_1, i_1; m_1, 1) &= b(i_1, i_1; m_1 + 1, 1) - b(i_1, i_1; m_1, 0) \text{ by (12),} \\ &= 0 \text{ by definition,} \end{aligned}$$

we do not require to use the polynomial $h(A_n)$ for which $r_{i_1} = m_1 - 1$. Hence in proving (16) we only use the $(e_1 + 1)(e_2 + 1) \cdots (e_t + 1) - (t + 1)$ polynomials $h(A_n)$ of hypothesis (b).

If $p \geq 2$, in (16) every equation is of the type,

$$c(j, k; s_j, 1)\sigma = 0, \quad (s_j = 1, 2, \cdots, e_j),$$

or by (12)

$$(26) \quad [(\lambda_j - \lambda_k)b(j, k; s_j, 1) + b(j, k; s_j + 1, 1)]\sigma = 0.$$

Since $\lambda_j \neq \lambda_k$, it follows that,

$$b(j, k; q, 1)\sigma = 0, \quad \text{if } b(j, k; q + 1, 1)\sigma = 0,$$

and, as $b(j, k; e_j + 1, 1) = 0$ by definition, that

$$b(j, k; s_j, 1)\sigma = 0, \quad (s_j = 1, 2, \cdots, e_j).$$

Accordingly, if $p \geq 2$, each letter c in (16) may be replaced by a letter b , so that (15) is true and Lemma 1 is proved.

If $p = 1$ in (16), the equation corresponding to (26) is

$$b(j, j; s_j + 1, 1) = 0, \quad (s_j = 1, 2, \cdots, e_j),$$

so that

$$(27) \quad b(j, j; s_j, 1) = 0, \quad (s_j = 2, 3, \cdots, e_j),$$

or every element in the first column of B_{jj} , except perhaps the first is zero.

If we now write $b(i, j)$ for the column vector, whose elements form the first column of the matrix B_{ij} , equation (15) becomes

$$(28) \quad b(i_1, i_2) b(i_2, i_3) \cdots b(i_p, i_1) = 0, \quad 2 \leq p \leq t.$$

The product on the left of (28) is a symbolic one and must be interpreted to mean (15). Consequently (28) is satisfied, if, and only if, for some value of $j \leq p$, $b(i_j, i_{j+1}) = 0$, $i_{p+1} = i_1$.

We proceed to prove

LEMMA 2. *If the t^2 vectors $b(i, j)$, $(i, j = 1, 2, \cdots, t)$, satisfy equations (28), there exists a permutation k_1, k_2, \cdots, k_t of the integers $1, 2, \cdots, t$, such that $b(k_r, k_s) = 0$, if r is greater than s .*

We shall prove this lemma by induction on t assuming that it is true for $t = 2, \cdots, m-1$. We note that the lemma is true when $m = 2$; for from the equation $b(12) b(21) = 0$ it follows that either $b(12) = 0$ or $b(21) = 0$ and that the lemma is true with $k_1 = 1, k_2 = 2$ or $k_1 = 2, k_2 = 1$.

Since the vectors $b(i, j)$, $(i, j = 1, 2, \cdots, m-1)$ satisfy (15) with $t = m-1$, by our induction assumption there exists a permutation $j_1, j_2, \cdots, j_{m-1}$ of the integers $1, 2, \cdots, m-1$, such that $b(j_r, j_s) = 0$, if $r > s$, $(r, s = 1, 2, \cdots, m-1)$. If we write

$$g(r, s) = b(j_r, j_s) \quad (r, s = 1, 2, \cdots, m; j_m = m),$$

we have

$$(29) \quad g(r, s) = 0, \quad r > s, \quad r \neq m,$$

and (15) becomes

$$(30) \quad g(i_1, i_2) g(i_2, i_3) \cdots g(i_p, i_1) = 0, \quad 2 \leq p \leq m.$$

If m does not occur in the set i_1, i_2, \cdots, i_p , (30) is satisfied by virtue of (29). Further, if $m = i_1$ and $i_j > i_{j+1}$ for some value of $j = 2, \cdots, p-1$, (30) is again satisfied, so that the equations (30), which are not satisfied because of (29), are all of the type

$$(31) \quad g(m, i_2) g(i_2, i_3) \cdots g(i_p, m), \quad i_2 < i_3 < \cdots < i_p; \quad 2 \leq p \leq m.$$

We now denote the equations (31), in which $g(j, m)$ appears, symbolically by

$$(32) \quad \{g(j, m)\} g(j, m) = 0, \quad (j = 1, 2, \cdots, m-1),$$

so that, if $g(j, m) \neq 0$, $\{g(j, m)\} = 0$. In (32) $\{g(j, m)\}$ represents a set of elements, each element being a product of one or more factors $g(r, s)$ and $\{g(j, m)\} = 0$ means that each element of the set is zero. In fact $\{g(j, m)\}$ is the set whose elements are

$$g(m, i_2) g(i_2, i_3) \cdots g(i_{p-1}, i_p) g(i_p, j), \quad i_2 < i_3 < \cdots < i_p < j; \quad 2 \leq p \leq j+1.$$

But the set of elements

$$g(m, i_2)g(i_2, i_3) \cdots g(i_{p-1}, i_p), \quad i_2 < i_3 < \cdots < i_p, \quad 2 \leq p \leq i_p + 1,$$

is simply the set $\{g(i_p, m)\}$. Consequently,

$$(33) \quad \{g(j, m)\} = g(m, j), \{g(1, m)\}g(1, j), \cdots, \{g(j-1, m)\}g(j-1, j), \\ (j=1, 2, \cdots, m-1).$$

We shall now show that for at least one value s , $1 \leq s \leq m$, $g(r, s) = 0$ for $(r=1, 2, \cdots, s-1, s+1, \cdots, m)$. If $\{g(j, m)\}$ is different from zero for all values of $j=1, 2, \cdots, m-1$, it follows from (32) that $g(j, m) = 0$, $(j=1, 2, \cdots, m-1)$ and that we may take $s=m$. Otherwise let $\{g(s, m)\} = 0$ but $\{g(j, m)\} \neq 0$, $j \leq s-1$; then by (33)

$$g(m, s) = g(1, s) = \cdots g(s-1, s) = 0$$

and, as by (29) $g(r, s) = 0$, when $r > s$ and $r \neq m$,

$$g(r, s) = 0, \quad (r=1, 2, \cdots, s-1, s+1, \cdots, m).$$

Accordingly there exists an integer s such that

$$(34) \quad b(j_r, j_s) = 0, \quad \text{if } r \neq s.$$

By our induction assumption there exists a permutation k_2, k_3, \cdots, k_m of the integers $j_1, j_2, \cdots, j_{s-1}, j_{s+1}, \cdots, j_m$, such that

$$(35) \quad b(k_r, k_f) = 0, \quad (r, f=2, \cdots, m, r > f).$$

If $j_s = k_1$, it follows from (34) and (35) that k_1, k_2, \cdots, k_m is a permutation of $1, 2, \cdots, m$ of such a nature that

$$b(k_r, k_f) = 0, \quad (r, f=1, 2, \cdots, m; r > f).$$

Accordingly our lemma is proved.

COROLLARY. *If $t=n$, that is, if all the latent roots of A are distinct the matrix $(b_{k_r k_s})$, $(r, s=1, 2, \cdots, n)$, is a triangle-matrix.*

This is an immediate consequence of the fact that each vector $b(k_r, k_s)$, being of dimension one, is the element $b_{k_r k_s}$.

If k_1, k_2, \cdots, k_t is the permutation of $1, 2, \cdots, t$ of Lemma 2 and

$$(36) \quad B_{k_r k_s} = D_{rs}; \quad (r, s=1, 2, \cdots, t),$$

the matrix $D = (D_{rs})$ is obtained from B_n by a permutation of the rows and the same permutation of the columns of B_n . But such a transformation of B_n is a similarity transformation,* so that there exists a non-singular matrix X_n satisfying the equation

$$(37) \quad X_n^{-1} B_n X_n = D.$$

By Lemma 2 and equation (27) all the elements in the first column of D are zero except perhaps the first. Hence

$$D = \begin{pmatrix} b_1 & \beta_1 \\ 0 & B_{n-1} \end{pmatrix},$$

where β_1 is a row vector of dimension $n-1$, 0 is the zero column vector of dimension $n-1$ and B_{n-1} a square matrix of order $n-1$. Similarly

$$X_n^{-1} A_n X_n = \begin{pmatrix} a_1 & \alpha_1 \\ 0 & A_{n-1} \end{pmatrix}$$

where α_1 is the vector $(1, 0, \dots, 0)$ of dimension $n-1$ and A_{n-1} is a square matrix of order $n-1$. Since

$$X_n^{-1} h(A_n) (A_n B_n - B_n A_n) X_n = \begin{pmatrix} 0 & \gamma \\ 0 & C_{n-1} \end{pmatrix},$$

where $C_{n-1} = h(A_{n-1}) (A_{n-1} B_{n-1} - B_{n-1} A_{n-1})$, if $h(A_n) (A_n B_n - B_n A_n)$ is nilpotent, so is $h(A_{n-1}) (A_{n-1} B_{n-1} - B_{n-1} A_{n-1})$. As a consequence of the nature of the matrix X_n in (35), A_{n-1} is still in canonical form; in fact

$$A_{n-1} = [M'_{k_1}, M_{k_2}, \dots, M_{k_t}],$$

where M'_{k_1} is the matrix of $e_{k_1} - 1$ rows and columns, obtained from M_{k_1} by removing the first row and the first column. Hence the polynomials of hypothesis (b), if defined for A_{n-1} instead of A_n would be $h(A_{n-1})$, where $h(A_n)$ is one of the polynomials (7) with r_{k_1} restricted to be at most $e_{k_1} - 1$. Accordingly by substituting A_{n-1} and B_{n-1} for A_n and B_n respectively and repeating our proof we show the existence of a non-singular $n-1$ rowed matrix Y , such that

$$(38) \quad Y^{-1} A_{n-1} Y = \begin{pmatrix} a_2 & \alpha_2 \\ 0 & A_{n-2} \end{pmatrix} \quad \text{and} \quad Y^{-1} B_{n-1} Y = \begin{pmatrix} b_2 & \beta_2 \\ 0 & B_{n-2} \end{pmatrix},$$

where α_2 and β_2 are row vectors of dimension $n-2$ and A_{n-2} and B_{n-2} square matrices of order $n-2$. Moreover, if

$$X_{n-1} = X_n \begin{pmatrix} 1 & 0 \\ 0 & Y \end{pmatrix},$$

*Turnbull and Aitken, *Canonical Matrices*, p. 11.

it follows from (37) and (38) that

$$X_{n-1}^{-1}A_nX_n = \begin{pmatrix} a_1 & a_{12} & \alpha_{13} \\ 0 & a_2 & \alpha_2 \\ 0 & 0 & A_{n-2} \end{pmatrix} \text{ and } X_{n-1}^{-1}B_nX_{n-1} = \begin{pmatrix} b_1 & b_{12} & \beta_{13} \\ 0 & b_2 & \beta_2 \\ 0 & 0 & B_{n-2} \end{pmatrix},$$

where the meaning of α_{13} and β_{13} is obvious. By repeating this process exactly $n-1$ times we find a non-singular matrix X_1 satisfying the equations

$$(39) \quad X_1^{-1}A_nX_1 = A_0 \text{ and } X_1^{-1}B_nX_1 = B_0,$$

where A_0 and B_0 are triangle-matrices. Moreover, since X_1 , in (39), is of the same type as X_n , A_0 and B_0 are derived from A_n and B_n respectively by a permutation of the rows and the same permutation of the columns. The matrix B_0 may be a triangle matrix of the most general type—that is, each element to the right of the leading diagonal may be different from zero but the matrix A_0 is not, since A_n is in canonical form. In fact in each row or column of A_0 there is at most one element, outside of the leading diagonal, which is different from zero.

Since A_n is the canonical form of A there exists a non-singular matrix Z such that, $Z^{-1}AZ = A_n$. If $Z^{-1}BZ = B_n$, then $h(A_n)(A_nB_n - B_nA_n)$ is nilpotent, if, and only if, $h(A)(AB - BA)$ is nilpotent. Moreover, if $W = ZX_1$, as a consequence of (39) we have

$$(40) \quad W^{-1}AW = A_0 \text{ and } W^{-1}BW = B_0.$$

Accordingly we have proved,

THEOREM I. *Let A be a square matrix of order n and let the elementary divisors of $A - \lambda E$ be*

$$(\lambda - \lambda_1)^{e_1}(\lambda - \lambda_2)^{e_2} \cdots (\lambda - \lambda_t)^{e_t}, \quad e_1 + e_2 + \cdots + e_t = n.$$

If A is not derogatory and if $h(A)(AB - BA)$ is nilpotent for each of the $(e_1 + 1)(e_2 + 1) \cdots (e_t + 1) - t - 1$ polynomials

$$h(A) = (A - \lambda_1 E)^{r_1}(A - \lambda_2 E)^{r_2} \cdots (A - \lambda_t E)^{r_t},$$

$$0 \leq r_i \leq e_i, \quad r_1 + r_2 + \cdots + r_t \leq n - 2,$$

then there exists a non-singular matrix W , satisfying (40), where B_0 is a triangle-matrix and A_0 is a triangle-matrix, derived from the classical canonical form of A by a permutation of the rows and the same permutation of the columns.

COROLLARY I. *If all the latent roots of A are distinct, a necessary and sufficient condition, that it be possible to reduce A to diagonal form and B to triangle form by the same similarity transformation, is hypothesis (b).*

For in this case the matrix A_0 is a diagonal matrix. It is interesting to compare this with the simpler but stronger condition, $AB - BA = 0$, for the possibility of a simultaneous reduction of A and B both to diagonal form.

COROLLARY II. *If A has a single elementary divisor, a necessary and sufficient condition, that it be possible to reduce A to canonical form and B to triangle form by the same similarity transformation, is hypothesis (b).*

For in this case A_0 is the same as A_n , since any permutation of the columns and the same permutation of the rows would destroy its triangle form. In this case the number of polynomials $h(A)$ of hypothesis (b) is a minimum namely $n - 1$, while in the previous case the number is a maximum, namely $2^n - n - 1$.

We now show by a simple example, that, if A is derogatory, hypothesis (b) is not sufficient to ensure the conclusion of Theorem 1.

$$\text{Let } A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Any polynomial $f(A)$ is of the form

$$\begin{pmatrix} \rho & 0 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & \rho + \sigma \end{pmatrix}$$

and accordingly,

$$f(A)(AB - BA) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\rho \\ \rho + \sigma & 0 & 0 \end{pmatrix}.$$

Since this last matrix is nilpotent for all values of ρ and σ , hypothesis (b) is certainly satisfied. Let $W^{-1}AW = A_0$ and $W^{-1}BW = B_0$, where A_0 and B_0 are triangle-matrices. Then $W^{-1}(\lambda A + \mu B)W = \lambda A_0 + \mu B_0$ identically in λ and μ , and in particular

$$(41) \quad |\lambda A + \mu B| \equiv |\lambda A_0 + \mu B_0|.$$

The determinant on the left of (41) has the value μ^3 and on the right the value $(\lambda + \omega_1\mu)\omega_2\omega_3\mu^2$ where $\omega_1, \omega_2, \omega_3$ are the three cube roots of unity. Hence (41) is not true and it is impossible to reduce A and B simultaneously to triangle form. Therefore, when A is derogatory, even if hypothesis (b) is

strengthened by replacing the finite number of polynomials $h(A)$ by all polynomials $f(A)$, it is not sufficient to ensure the simultaneous reduction of A and B to triangle form.

If A is derogatory, but for some value of λ , $A + \lambda B = C$ is not derogatory, we may apply Theorem 1 to the matrices C and B . In hypothesis (b), $h(A)$ must be replaced by $h(C)$ and the nilpotent polynomials by $h(C)(CB - BC)$. The matrix $h(C)CB$ is certainly a polynomial in A and B , say $f(A, B)$, and $h(C)BC$ a second such polynomial $\phi(A, B)$. Moreover, if x and y are indeterminates

$$(42) \quad f(x, y) - \phi(x, y) \equiv 0.$$

Hence, if for every pair of polynomials f and ϕ , which satisfy (42), $f(A, B) - \phi(A, B)$ is nilpotent, it is possible to reduce C and B , and therefore A , to triangle form by the same similarity transformation. It seems probable that a similar result holds even when every matrix of the pencil is derogatory but as yet we have been unable to prove it.

As a consequence of Theorem 1, we have

THEOREM 2. *If A is not derogatory a necessary and sufficient condition that the latent roots of $f(A, B)$ be $f(\lambda_i, \mu_i)$, for every polynomial $f(A, B)$, where λ_i and μ_i are the latent roots of A and B respectively, is that hypothesis (b) be satisfied.**

For, if (b) is true, A and B can be reduced simultaneously to triangle form and hence the latent roots of $f(A, B)$ are $f(\lambda_i, \mu_i)$. Conversely if the latent roots of $f(A, B)$ are $f(\lambda_i, \mu_i)$, the latent roots of $h(A)(AB - BA)$ are all zero, so that $h(A)(AB - BA)$ is nilpotent and (b) is satisfied.

As a triangle-matrix is the canonical form of a matrix under unitary transformation † it is to be expected that a theorem similar to Theorem I should hold, if unitary transformations are employed instead of similarity transformations. This is in fact the case. Since the matrix W in (40) is non-singular there exists a triangle-matrix T such that $WT = U$ is a unitary matrix.‡ We have therefore from (40)

* This problem has also been considered by G. S. Bruton, "Certain aspects of the theory of equations for a pair of matrices," and M. H. Ingraham, "A study of related pairs of square matrices." Abstracts of these papers appear in the *Bulletin of the American Mathematical Society*, vol. 38 (1932), p. 633. N. H. McCoy in his paper "Quasi-commutative matrices," *Transactions of the American Mathematical Society*, vol. 36 (April, 1934), shows that if A and B are quasi-commutative the latent roots of $f(A, B)$ are $f(\lambda_i, \mu_i)$.

† Turnbull and Aitken, *op. cit.*, p. 94.

‡ Turnbull and Aitken, *op. cit.*, p. 96. Schmidt's Theorem.

$$\begin{aligned}T^{-1}W^{-1}AWT &= U^*AU = T^{-1}A_0T = T_1, \\T^{-1}W^{-1}BWT &= U^*BU = T^{-1}B_0T = T_2,\end{aligned}$$

where, since the inverse of a triangle-matrix is a triangle-matrix, T_1 and T_2 are triangle-matrices.

Hence we have,

THEOREM 3. *If A is not derogatory, a necessary and sufficient condition, that it be possible to reduce A and B to triangle form, both by the same unitary transformation, is that hypothesis (b) be satisfied.*

THE JOHNS HOPKINS UNIVERSITY.

SINGULARITIES OF ANALYTIC VECTOR FUNCTIONS.

By SI-PING CHEO.

1. *Preliminary considerations.* There are many methods of extending the theory of ordinary analytic functions to three dimensional space or better of constructing a theory of functions of three variables which would be analogous to the theory of ordinary analytic functions. For example, expansions in power series, conformal representation, Cauchy's method based on monogeneity, etc. are all capable of leading to various extensions of the theory of ordinary analytic functions. The theory we have in mind here is based on the generalization of the Cauchy-Riemann differential equations.

Definition of analytic vector functions. If we have three functions X, Y, Z of three real variables x, y, z which are Cartesian coördinates of a point in space, if all the partial derivatives of the first order exist and are continuous in a certain region R , and if the following conditions,

$$(1.1) \quad \begin{aligned} \operatorname{div} \vec{\Phi} &= \operatorname{div} (X\vec{i} + Y\vec{j} + Z\vec{k}) = 0 & \vec{i}, \vec{j}, \vec{k}, \text{ unit vectors per-} \\ \operatorname{curl} \vec{\Phi} &= \operatorname{curl} (X\vec{i} + Y\vec{j} + Z\vec{k}) = 0 & \text{pendicular to each other,} \end{aligned}$$

are satisfied in R , then we shall say the vector function, $\vec{\Phi}$ is analytic throughout R .

The above set of equations has been considered as a generalization of the set of the Cauchy-Riemann differential equations.*

By the fundamental theorems of vector calculus, we notice that from the first equation of (1.1) $\vec{\Phi}$ must be the curl of a vector function $\vec{\Psi}$ (say), and from $\operatorname{curl} \vec{\Phi} = 0$, $\vec{\Phi}$ must be the gradient of a scalar function H (say); thus we obtain the following relation:

$$(1.2) \quad \operatorname{curl} \vec{\Psi} = \operatorname{grad} H.$$

From the above relation, we can easily see $\nabla^2 H = 0$ and $\operatorname{grad} \operatorname{div} \vec{\Psi} = \nabla^2 \vec{\Psi}$, where ∇^2 denotes the Laplace Operator. In fact, we could state the following two lemmas:

* G. Y. Rainich, "Analytic functions and mathematical physics," *Bulletin of the American Mathematical Society* (October, 1931).

LEMMA 1. A necessary and sufficient condition for a vector function, $\vec{\Phi} = \text{grad } H$, to be analytic is that H must be harmonic.*

LEMMA 2. A necessary and sufficient condition for a vector function, $\vec{\Phi} = \text{curl } \vec{\Psi}$ to be analytic is

$$(1.3) \quad \text{grad div } \vec{\Psi} = \nabla^2 \vec{\Psi}.$$

The above two lemmas suggest us that we may have two ways of obtaining analytic vector functions from *harmonic functions*. The first consists simply in taking the gradient of a harmonic function; a function obtained in this way we shall call a *gradient function*. The second consists in going through the following steps:

- 1) Replacing x, y, z in a harmonic function $H(x, y, z)$ by $x_2 - x_1, y_2 - y_1, z_2 - z_1$, respectively.
- 2) Integrating $H(x_2 - x_1, y_2 - y_1, z_2 - z_1)$ along a close curve C_1 with respect to x_1, y_1, z_1 , that is, taking $\int_{C_1} H(x_2 - x_1, y_2 - y_1, z_2 - z_1) \vec{ds}_1$ where $\vec{ds}_1 = dx_1 \vec{i} + dy_1 \vec{j} + dz_1 \vec{k}$ is the curve element of C_1 .
- 3) Taking the curl with respect to x_2, y_2, z_2 of $\int_{C_1} H \vec{ds}_1$, that is, taking $\text{curl}_2 \int_{C_1} H(x_2 - x_1, y_2 - y_1, z_2 - z_1) \vec{ds}_1$.

We shall call this process the Ω -process; and the functions which are obtained by Ω -process will be called Ω -functions. Now we can state the following theorem:

THEOREM 1. Ω -functions are always analytic.

Without any difficulty, this theorem may be proved rigorously; and it is quite obvious from the view-point of mathematical physics.†

2. *Singularities.* An analytic vector function in three dimensional space may have isolated singular points, and it may also have isolated singular curves. The definitions of these singularities seem to be very natural, and are given as follows: A point is said to be an *isolated singular point* of a given analytic vector function, provided that this function is not analytic at

* In order to express ourselves briefly, we shall define a harmonic function in the following way: A function which possesses all continuous partial derivatives of the first and second orders and satisfies Laplace's equation will be called a harmonic function.

† See, for example, Livens, *Theory of Electricity* (1918), p. 356.

that point, but at all points in the neighborhood of this point, the function of analytic. A curve is said to be an *isolated singular curve* of a given analytic vector function, provided that this function is not analytic at any of the points of the curve, but at all points in the neighborhood of the curve, the function is analytic.

An isolated singular point and an isolated singular curve will be called briefly a *singular point* and a *singular curve*, respectively.

We want to investigate now the singularities of the two kinds of analytic vector functions introduced in the preceding section.

If the harmonic function H which has been used in the formation of a gradient function has a singular point* at (a, b, c) , then the gradient function will also have a singularity at that point. Furthermore, we notice that the operator gradient does not introduce any new singularity. Hence, we can state the following theorem:

THEOREM 2. *A gradient function possesses the same singularities as those of the corresponding harmonic function.*

Let us now investigate the singularities of Ω -functions. Consider the vector function,

$$\vec{\Phi} = \text{curl}_2 \int_{C_1} (1/\gamma_{21}) d\vec{s}_1$$

where $\gamma_{21}^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2$. It is well-known the function $1/\gamma_{21}$ is single-valued and harmonic everywhere in space except at the origin. Therefore $\vec{\Phi}$ is analytic everywhere in space except when $x_2 = x_1$, $y_2 = y_1$, $z_2 = z_1$; that is to say, $\vec{\Phi}$ is not analytic at every point of the curve C_1 . It will be seen in the next section that C_1 is the singular curve of $\vec{\Phi}$. In general, if a single-valued harmonic function possesses a singular point at (a, b, c) , then the corresponding Ω -function is analytic everywhere in space, except at all the points of $C_1(\vec{a}i + \vec{b}j + \vec{c}k)$, which is obtained from C_1 by translating it through the vector, $\vec{a}i + \vec{b}j + \vec{c}k$. In fact, we could state the following theorem:

THEOREM 3. *If a single-valued harmonic function possesses n singular points at (a_1, b_1, c_1) , (a_2, b_2, c_2) , \dots (a_n, b_n, c_n) , then the corresponding Ω -function will be defined and analytic everywhere in space except on the points of the n congruent curves $C_1(\vec{a}_1i + \vec{b}_1j + \vec{c}_1k)$, $C_1(\vec{a}_2i + \vec{b}_2j + \vec{c}_2k)$, \dots , $C_1(\vec{a}_ni + \vec{b}_nj + \vec{c}_nk)$, which are obtained from C_1 by translating it through the following vectors: $\vec{a}_1i + \vec{b}_1j + \vec{c}_1k$, $\vec{a}_2i + \vec{b}_2j + \vec{c}_2k$, \dots $\vec{a}_ni + \vec{b}_nj + \vec{c}_nk$, respectively.*

* That is to say, H is harmonic everywhere in space except at (a, b, c) .

The Ω -process breaks down for the points which lie on the curves, $C_1 \xrightarrow{\rightarrow} \xrightarrow{\rightarrow} \xrightarrow{\rightarrow} C_2 \xrightarrow{\rightarrow} \xrightarrow{\rightarrow} \xrightarrow{\rightarrow} \dots, C_1 \xrightarrow{\rightarrow} \xrightarrow{\rightarrow} \xrightarrow{\rightarrow} C_1 \xrightarrow{\rightarrow} \xrightarrow{\rightarrow} \xrightarrow{\rightarrow} C_1$. Whether or not it is possible to assign values to the function at the points of these curves in such a way as to make the vector function analytic on these curves, the next section will tell us.

3. *Residues of analytic vector functions.* The first equation of (1.1) implies the vanishing of a surface integral,

$$(3.1) \quad \int \int_S (Xl + Ym + Zn) d\sigma,$$

where S is a surface lying within the region R and which can be contracted to a point without going outside of R ; and l, m, n are the direction cosines of the normal * to S . This can be seen by Gauss' Theorem, which states:

$$\int \int \int_V (\partial X / \partial x + \partial Y / \partial y + \partial Z / \partial z) d\tau = \int \int_S (Xl + Ym + Zn) d\sigma,$$

V being the volume bounded by S .

In (1.1), $\text{curl } \vec{\Phi} = 0$ is the condition for the vanishing of a curve integral:

$$(3.2) \quad \int_C (Xdx + Ydy + Zdz),$$

where the curve C lies entirely in R , and can be contracted to a point without going outside R . In this case the proof is based on the following identity:

$$\begin{aligned} \int \int_S \left\{ (\partial Z / \partial y - \partial Y / \partial z)l + (\partial X / \partial z - \partial Z / \partial x)m + (\partial Y / \partial x - \partial X / \partial y)n \right\} d\sigma \\ = \int_C (Xdx + Ydy + Zdz), \end{aligned}$$

S being a surface bounded by C . The above relation is known as Stokes' Theorem.

It may be the case that we can not contract S , and C to a point without going outside R , then the surface integral (3.1) and the curve integral (3.2) may have values different from zero, say K_s and K_c , respectively. We shall call $(1/4\pi)K_s$ the surface-residue, and $(1/4\pi)K_c$ the curve-residue of the vector function, $\vec{\Phi}$, given by S and C respectively.

Suppose $\vec{\Phi}$ has an isolated singular point. This point must be considered as not belonging to R ; therefore, the surface S enclosing this point can not

* We shall assume that the normal to be directed inward.

be contracted to a point without going outside R . In this case, the surface residue might be different from zero. We notice that this surface residue is independent from the surface which encloses the singular point. In fact, two surfaces which enclose the same singular point and no other singularities can be transformed, one from the other, without going outside the region in which the vector function is analytic; therefore, they will give the same residue. We shall call this value the *Surface-Residue* of that function with respect to the singular point.

What could prevent the integral (3.2) from being zero is the existence of a closed singular curve of the vector function, $\vec{\Phi}$. In case a curve links the singular curve, it can not be contracted to a point without going outside R . Two curves which can be transformed one into the other without going outside R give the same residue, regardless of sign. In particular, two curves each of which links a given singular curve once can be so transformed into each other; therefore, they give the same residue. We shall call the residue given by a curve which links *once* with a singular curve of a vector function, the *Curve-Residue* of the vector function with respect to the singular curve.

Summarizing the above considerations and using the notations of vector calculus, we can define these two kinds of residues of analytic functions as follows:

If S is a closed surface lying in the region of analyticity of a vector function $\vec{\Phi}$ but enclosing giving singularities of that function, then the surface integral $(1/4\pi) \int_S \vec{\Phi} \cdot \vec{n} d\sigma$ will be called the surface residue of $\vec{\Phi}$ with respect to the singularities, where \vec{n} the unit normal of $d\sigma$ directs toward the interior of S , $d\sigma$ in the element of S , and the dot (\cdot) is used as a sign of scalar product.

The curve residue of $\vec{\Phi}$ with respect to its singularities will be defined as the curve integral $(1/4\pi) \int_C \vec{\Phi} \cdot \vec{ds}$, where C is a closed curve lying entirely in R and links only *once* with each of the singularities, \vec{ds} in its "positive sense" denotes the element of C .

Suppose now a gradient function $\vec{\Phi}$ having a singular point $P_0(x_0, y_0, z_0)$ in a certain region R and taking the following form:

$$\text{grad } \frac{K_s}{\gamma_0}, \quad K_s = \text{constant} \\ \gamma_0^2 = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2.$$

Then, by definition, the surface residue of the gradient function with regard to P_0 will be:

$$\frac{1}{4\pi} \int_S \operatorname{grad} \frac{K_s}{\gamma_0} \cdot \vec{n} d\sigma = \frac{1}{4\pi} \int_S \frac{\partial}{\partial n} \frac{K_s}{\gamma_0} d\sigma = K_s$$

as we have seen that it is true in the theory of potentials. In fact, we can state:

THEOREM 4. *The surface-residue of a gradient function with regard to a certain singular point in a certain region is a constant.**

Suppose that an Ω -function has a singular curve C_1 and take the following form:

$$\vec{\Phi} = \operatorname{curl}_2 \int_{C_1} \frac{K_c}{\gamma_{21}} d\vec{s}_1,$$

where K_c is a constant. According to the definition, the curve-residue of $\vec{\Phi}$ is:

$$\begin{aligned} \frac{1}{4\pi} \int_{C_2} \vec{\Phi} \cdot d\vec{s}_2 &= \frac{1}{4\pi} \int_{C_2} \operatorname{curl}_2 \int_{C_1} \frac{K_c}{\gamma_{21}} d\vec{s}_1 \cdot d\vec{s}_2, \quad d\vec{s}_2 = dx_2 \vec{i} + dy_2 \vec{j} + dz_2 \vec{k} \\ &= \frac{K_c}{4\pi} \int_{C_2} \int_{C_1} \operatorname{curl}_2 \frac{1}{\gamma_{21}} d\vec{s}_1 \cdot d\vec{s}_2 \\ &= \frac{K_c}{4\pi} \int_{C_2} \int_{C_1} \frac{(x_2 - x_1)(dy_1 dz_2 - dz_1 dy_2)}{\gamma_{21}^3} \\ &\quad + \frac{(y_2 - y_1)(dz_1 dx_2 - dx_1 dz_2) + (z_2 - z_1)(dx_1 dy_2 - dy_1 dx_2)}{\gamma_{21}^3} \\ &= \frac{K_c}{4\pi} \int_{C_2} \int_{C_1} \frac{1}{\gamma_{21}^3} \begin{vmatrix} x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ dx_1 & dy_1 & dz_1 \\ dx_2 & dy_2 & dz_2 \end{vmatrix} \\ &= M K_c, \end{aligned}$$

where M , an integer,† denotes the number of times for which C_1 and C_2 are

* A gradient function of the general form, $\vec{\Phi} = \operatorname{grad} H$ having a singular point at $P_0(x_0, y_0, z_0)$ may be developed around P_0 in a power series of the form,

$$\sum_{n=0}^{\infty} (h_n / \gamma_0^{2n+1}),$$

where h_n is a homogeneous, harmonic function of n -th degree. We can verify that the residue of $\vec{\Phi}$ is h_0 which is a constant.

† *Gauss Werk*, Band V (1877), p. 605. See also Boeddicker, *Gauss'schen Theorie der Verschlingungen*, Stuttgart (1876); Urysohn, "Sur les multiplicités Cantorienes," *Fundamenta Mathematicae*, vols. 7-8 (1925-26).

linked together. By the definition of curve-residue, C_1 and C_2 are linked together only *once*, therefore M is here equal to the unit. Hence, the curve-residue of the above Ω -function with respect to C_1 is K_c . In fact, we can state the following theorem:

THEOREM 5. *The curve-residue of Ω -function with respect to a certain singular curve in a certain region is a constant.**

We notice that the following integral:

$$\frac{1}{4\pi} \int_{C_1} \text{curl}_1 \int_{C_2} \frac{K_c}{\gamma_{21}} \vec{ds}_2 \cdot \vec{ds}_1$$

which is equal to

$$-\frac{K_c}{4\pi} \int_{C_1} \int_{C_2} \begin{vmatrix} x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ dx_2 & dy_2 & dz_2 \\ dx_1 & dy_1 & dz_1 \end{vmatrix}$$

is K_c also. Hence, we may state:

THEOREM 6. *The curve-residue the Ω -function, $\text{curl}_2 \int_{C_1} (1/\gamma_{21}) \vec{ds}_1$, with respect to C_1 is identical to that of the Ω -function, $\text{curl}_1 \int_{C_2} (1/\gamma_{21}) \vec{ds}_2$, with respect to C_2 .*

There are many problems regarding analytic vector functions remaining unsolved. It would be very interesting to generalize all the considerations in the previous discussions; that is to say, to increase the number of dimensions, and to find the relationships between analytic functions and their different kinds of isolated singularities.

UNIVERSITY OF MICHIGAN.

* An Ω -function of the general form, $\vec{\Phi} = \text{curl}_2 \int_{C_1} H \vec{ds}_1$, having a singular curve C_1 may be developed "around C_1 " into power series of the form:

$$\sum_{n=0}^{\infty} \text{curl}_2 \int_{C_1} (h_n / \gamma_{21}^{2n+1}) \vec{ds}_1$$

where h_n is a homogeneous, harmonic function of degree n . We can verify that the curve-residue of this Ω -function is h_0 which is a constant.

THE STRUCTURE OF A COMPACT CONNECTED GROUP.

By E. R. VAN KAMPEN.

I. In a recent paper Pontrjagin proved implicitly the following theorem:

*If U is any nucleus of a compact group F , then U contains a closed invariant subgroup H of F , such that F/H is a (not necessarily connected) Lie group.**

Applying this theorem to a sequence of nuclei of F , converging to the identity element 1 of F , we can construct a decreasing sequence of closed invariant subgroups H_n , also converging to 1, such that all factor groups $F_n = F/H_n$ are Lie groups. If $m > n$, the group $H_n/H_m = H_{nm}$ is a subgroup of $F/H_m = F_m$ and then F_n can be identified with the factor group F_m/H_{nm} .†

It can be proved very easily that F is uniquely determined by the sequence of groups F_n and the identities $F_n = F_m/H_{nm}$, $m > n$. It is even possible to construct F if a sequence of groups F_n and identities $F_n = F_m/H_{nm}$, $m > n$, is given, provided these identities satisfy an obvious transitive law.‡ However we will not need to construct a group by this method.

We consider connected groups F only. Then all groups F_n are connected also, and we can use the known structural properties of compact connected Lie groups § to find structural properties of F . By means of the relations $F_n = F_m/H_{nm}$ we establish in II relations between the structural elements of all groups F_n . Then a simple limiting process (described in III) allows to draw conclusions about F (IV). In V we make the analogous conclusions for certain finite covering groups of the groups F_n .

The results of these sections will be found in Theorems 1 and 2. The

* L. Pontrjagin, "Sur les groupes topologiques compacts," *Comptes Rendus*, vol. 198 (1934), p. 238. An explicit formulation and proof will be found in a paper by E. R. van Kampen to appear shortly in the *Annals of Mathematics*. A nucleus is an open set containing the identity element. Compare: E. R. van Kampen, "Locally bicomact abelian groups," *Annals of Mathematics*, vol. 36 (1935), no. 2, I, 2.

† Whenever no contradiction arises as a consequence, we do not hesitate to call simply isomorphic groups identical. This frequently leads to a considerable simplification in notation and language.

‡ Compare the paper by Pontrjagin mentioned above.

§ See E. Cartan, "La théorie des groupes finis et continus," *Mém. d. Sc. Math.*, Fasc. 42, p. 42. We suppose that the reader is acquainted with his results.

difference between the general case and the case of a Lie group is not greater than the minimum that was to be expected. Certain finite abelian groups have to be replaced by 0-dimensional compact abelian groups and certain finite direct products of (locally) simple groups by countable direct products.

In the remaining three sections we discuss the structure of the 0-dimensional groups occurring, the behavior of F as regards local connectedness, and a generalized idea of covering space naturally arising as a consequence of the relations between D , $D/B = F$ and F/A . (See Theorems 1 and 2.)

For the common part of two groups we write $A.B$. For the direct product of A, B, \dots we use the notation $[A + B + \dots]$. The symbol (A, B) denotes the group generated by A and B . If a group A is a covering group of another group B we call the groups locally isomorphic. In that case the multiple isomorphism of A and B is such that for sufficiently small nuclei it is one-to-one and bicontinuous.

II. The compact connected Lie group F_n , ($n = 1, 2, \dots$), contains a number of (locally) simple invariant subgroups. The semi-simple subgroup S_n of F_n generated by all these simple groups has a finite group A_n in common with the centrum C_n of F_n . The factor group F_n/A_n is the direct product of C_n/A_n and S_n/A_n ; and S_n/A_n is the direct product of simple Lie groups, each with degenerate centrum (consisting of 1 only).

Comparing F_m with $F_n = F_m/H_{nm}$, we see that H_{nm} must either contain any of the simple subgroups of F_m or meet it in at most a finite number of centrum elements. As a consequence we can find a (finite or infinite) sequence of simple Lie groups $\bar{S}^{(1)}, \bar{S}^{(2)}, \dots$ each with degenerate centrum, and a non-decreasing sequence of integers p_1, p_2, \dots such that the simple groups occurring in S_n/A_n are simply isomorphic with $\bar{S}^{(1)}, \dots, \bar{S}^{(p_n)}$.

Of course we may suppose that the subgroup $S_m^{(l)}$ of F_m corresponding to $\bar{S}^{(l)}$ has as image under the transformation defined by $F_n = F_m/H_{nm}$ the subgroup $S_n^{(l)}$ of F_n corresponding to the same $\bar{S}^{(l)}$. Here $S_n^{(l)}$ can be taken as the identity element of F_n , whenever $l > p_n$.

The image of the semi-simple subgroup S_m of F_m under the same transformation of F_m into F_n must be the corresponding subgroup S_n of F_n . For this is locally true and S_n is in F_n determined by its infinitesimal transformations.

But also the image of the centrum C_m of F_m is equal to the centrum C_n of F_n . The image C_n^* of C_m is obviously contained in C_n . The factor group of C_n^* in F_n is simply isomorphic with the factor group of $H_{nm}/(C_m \cdot H_{nm})$ in $F_m/C_m = S_m/A_m$. As any factor group of $S_m/A_m = [\bar{S}^{(1)} + \dots + \bar{S}^{(p_m)}]$

has a degenerate centrum it follows immediately that F_n/C_n^* has a degenerate centrum, so that $C_n^* = C_n$.

Applying this reasoning to S_m and its centrum A_m instead of F_m and its centrum C_m we find that also A_n is the image of A_m under the transformation determined by $F_n = F_m/H_{nm}$.

III. Any invariant subgroup G_m of F_n determines uniquely a largest invariant subgroup G'_n of F , such that G'_n is transformed into G_n under the transformation determined by $F_n = F/H_n$. Suppose G_n is defined for all n ; then the common part G of all groups G'_n is a well defined closed invariant subgroup of F . Suppose the image of G_m under the transformation defined by $F_n = F_m/H_{nm}$ is contained in G_n ; then G'_n decreases with increasing n , and the image of G under the transformation defined by $F_n = F/H_n$ is contained in G_n .

Finally, suppose that the image in F_n of the group G_m in F_m is equal to G_n . Then the image in F_n of the subgroup G'_m of F , under the transformation defined by $F_n = F/H_n$ ($m > n$) is equal to G_n , so the image of G in F_n is also G_n . But then $G/(H_n \cdot G) = G_n$ and $H_n \cdot G$ is arbitrarily small in G , so that G is approximated by the groups G_n , in the way described in I for F and F_n .

IV. We apply this on the system of subgroups defined in II, finding invariant subgroups $S^{(l)}$, S , C , A of F , corresponding to the subgroups $S_n^{(l)}$, S_n , C_n , A_n of F_n .

The groups $S_n^{(l)}$ are for $p_n > l$, locally isomorphic with $\bar{S}^{(l)}$, so there are only a finite number of possibilities for the structure of $S_n^{(l)}$ and for sufficiently large n all groups $S_n^{(l)}$ (l fixed), must be simply isomorphic. But then they are simply isomorphic with $S^{(l)}$. So $S^{(l)}$ is a compact simple Lie group, locally isomorphic with $\bar{S}^{(l)}$.

It is clear that S is contained in the group generated by $S^{(1)}, \dots, S^{(p_n)}$ and H_n and on the other hand that S contains all $\bar{S}^{(l)}$. So S is equal to the group generated by all $S^{(l)}$.

We can see directly that the common part of all groups C'_n is the centrum of F , so C is the centrum of F . Applying this on S we see that A is the centrum of S .

As each image A_n of A is finite, A must be 0-dimensional; as A_n is the common part of S_n and C_n , A is the common part of S and C ; as S_n and C_n together generate F_n , so S and C generate F .

Under the transformation defined by $F_n = F_m/H_{nm}$ the image of any co-set of A_m is a co-set of A_n , so there is an invariant subgroup of F_m/A_m of

which the factor group is F_n/A_n . It can be verified immediately that F_n/A_n can be obtained from F/A by taking the factor group of $(H_n, A)/A$. As F_n/A_n is the direct product of C_n/A_n and $\bar{S}^{(1)}, \dots, \bar{S}^{(p_n)}$ we can also obtain F_n/A_n as the factor group of an arbitrarily small invariant subgroup of the direct product of C/A and all $\bar{S}^{(l)}$. So because any compact group is uniquely determined by its approximating groups, F/A must be the direct product of C/A and all groups $\bar{S}^{(l)}$.

All these results can be combined in the following theorem:

THEOREM 1. *Suppose F is a compact connected group, $S^{(l)}$, $l = 1, 2, \dots$, are all the (locally) simple (compact) Lie groups invariant in F , S is the group generated by all $S^{(l)}$ and C is the centrum of F . Then S and C generate F , and have in common the 0-dimensional centrum A of S . The factor group F/A is simply isomorphic with the direct product of C/A and all groups $\bar{S}^{(l)}$, where $\bar{S}^{(l)}$ is the simple group with degenerate centrum locally isomorphic with $S^{(l)}$.*

COROLLARY. *If a compact connected group F has a degenerate centrum, then it is the direct product of a collection of simple Lie groups.*

V. For each group F_n we define a covering group D_n in the following way: An element of D_n is an oriented arc α in F_n beginning in the identity element of F_n . Two such elements α and β are called equal if they have the same endpoint and the simple closed curve $\alpha\beta^{-1}$ is isotopic with a curve in the maximal connected subgroup K_n of the centrum C_n of F_n . The product $\alpha\beta$ is defined as the arc $\alpha\beta'$, where β' is obtained from β by left multiplication with the endpoint of α .

As A_n is a finite group, D_n can also be defined as covering group of F_n/A_n ; the simple closed curves of F_n/A_n , corresponding to the identity element of D_n are then isotopic with curves in C_n/A_n , but not with arbitrary such curves. Anyway we can see that D_n is the direct product of simply connected simple groups $\bar{S}^{(1)}, \dots, \bar{S}^{(p_n)}$ (locally isomorphic with $S^{(1)}, \dots, S^{(p_n)}$) and a group L_n , locally isomorphic with C_n/A_n (or with K_n). As simple closed curves in K_n correspond to the identity element of D_n it follows now that K_n and L_n are simply isomorphic and that D_n is compact. So D_n is a finite covering group of F_n , and it must have a finite centrum subgroup B_n , such that $D_n/B_n = F_n$. The groups B_n and L_n can only have the identity element in common.

The transformation of F_m into F_n defined by $F_n = F_m/H_{nm}$ can be used to define a transformation of D_m into D_n : As image in D_n of an element α

of D_m , we take the image in F_n of the arc α . The transformation so defined is independent of the particular arc chosen to determine the element of D_m , for the image of a simple closed curve in F_m , isotopic with a curve in K_m is a simple closed curve in F_n , isotopic with a curve in K_n . As apparently the image of a product is equal to the product of the images the transformation is a multiple isomorphism. So D_m has a certain invariant subgroup D_{nm} , such that $D_n = D_m/D_{nm}$ and that the resulting transformation of D_m into D_n is the one we are considering.

We can find a nucleus U of D_m , for which the transformation into F_m defined by $D_m/B_m = F_m$ is a homeomorphism and such that the same is true for the image V of U in D_n . Then the transformation of U into V defined by $D_n = D_m/D_{nm}$ is the same as the transformation of the corresponding nuclei in F_m and F_n defined by $F_n = F_m/H_{nm}$. It follows immediately that the transformation of D_m into D_n has the following properties:

1. The subgroup of D_m corresponding to $\bar{S}^{(l)}$ is transformed into the subgroup of D_n corresponding to $\bar{S}^{(l)}$. If $l > p_n$ the last subgroup is the identity element of D_n . If $l \leq p_n$ the correspondence between these two groups is a simple isomorphism.

2. The transformation of the subgroup L_m of D_m into the subgroup L_n of D_n can be obtained by applying in succession the simple isomorphism of L_m and K_m , the transformation of K_m into K_n defined by $F_n = F_m/H_{nm}$ and the simple isomorphism of K_n and L_n .

From 1 and 2 it follows that the groups D_n can be considered as approximating groups for a group D defined as the direct product of all groups $\bar{S}^{(l)}$ and a group L simply isomorphic with the maximal connected subgroup K of the centrum C of F .

The image in D_n of the subgroup B_m of D_m is continued in B_n . For an element of B_m is a simple closed curve α in F_m ; the image of α is a simple closed curve in F_n , that means an element of B_n .

So according to III the groups B_n determine an invariant subgroup B of D . As the image of B in D_n is part of B_n it is finite, so B is 0-dimensional. As L_n and the image of B in D_n have only the identity element in common, so L and B have only the identity element in common.

If B'_n is the subgroup of D corresponding to the subgroup B_n of D_n (compare III for G'_n in F corresponding to G_n in F_n), then the factor group of B'_n/B in D/B is simply isomorphic with D/B'_n . But D/B'_n is simply isomorphic with $D_n/B_n = F_n$. At the same time B'_n/B is arbitrarily small in D/B because B is the common part of all B'_n . So the group D/B is

approximated (in the sense of I) by the sequence of groups F_n . As any compact group is uniquely determined by its approximating sequence, it follows that $D/B = F$. So we have proved:

THEOREM 2. *Suppose for the group F of Theorem 1, K is the maximal connected subgroup of the centrum, $\tilde{S}^{(1)}$ is the simply connected group locally isomorphic with $S^{(1)}$ and D is the direct product of all $\tilde{S}^{(1)}$ and a group L simply isomorphic with K . Then D has a 0-dimensional invariant subgroup B meeting L only in the identity element and such that $F = D/B$. The subgroup B is uniquely determined up to automorphisms of D .*

Remark: The image of B_m in D_n is in general not equal to B_n (as might be expected after the considerations in III) but only contained in B_n . The reason is that it may be impossible to obtain D_n from D and F_n from $F = D/B$ using one invariant subgroup of D . Once the construction of D is completed, we can easily find a new sequence of groups approximating F and such that the image of the group corresponding to B_m is equal to the group corresponding to B_n . We have to find invariant subgroups T_n of D such that $D/T_n = D_n$ and then use the invariant subgroups $(T_n, B)/B$ to define the new factorgroups of F .

V. An investigation of the character of the two 0-dimensional abelian groups A and B shows that while B is the most general type, A is of very simple structure: A direct sum of finite cyclic groups.

The centrum of D is the direct product of the connected abelian group L and the centrum M of the direct product of all groups $\tilde{S}^{(1)}$. Investigations of Cartan* show that M is an arbitrary (compact) direct product of finite cyclic groups. As each co-set of L in the centrum of D has with B at most one element in common and has with M exactly one element in common, it follows that B is simply isomorphic with an arbitrary closed subgroup of M . As arbitrary closed subgroup of an arbitrary compact direct product of finite cyclic groups, B is an arbitrary 0-dimensional abelian group.†

On the other hand, A is the centrum of S and S is simply isomorphic

* See E. Cartan, *loc. cit.*, p. 41.

† The theorems on 0-dimensional abelian groups here used are readily verified by reducing them to corresponding theorems for their character groups. See L. Pontrjagin, *Annals of Mathematics*, vol. 35 (1934), pp. 361-388 and E. R. van Kampen, *Annals of Mathematics*, vol. 36 (1935), no. 2. The character group of $B(A)$ is an arbitrary factor group (subgroup) of a discrete countable direct product of finite cyclic groups. And it can be verified immediately that the character group of B is an arbitrary countable abelian group without elements of infinite order, while the character group of A is a discrete countable direct product of finite cyclic groups.

with the factor group of an arbitrary subgroup of M in the direct product of all $\bar{S}^{(i)}$. So A is the factor group of an arbitrary closed subgroup of M . As such it is itself a (compact) direct product of finite cyclic groups.*

VI. The direct sum of all groups $\bar{S}^{(i)}$ is locally connected, so its image S is also locally connected. If K is also locally connected, then D and its image F are locally connected. On the other hand, if F is locally connected, then its image F/A is also locally connected and so C/A is locally connected. So it is to be expected that F and some group connected with its centrum will be locally connected or not locally connected at the same time. It is quite easy to verify that F can be locally connected, while K is not locally connected. The following theorem shows the precise relationship:

THEOREM 3. *A compact connected group F is locally connected if and only if the group C/A (defined in Theorem 1) is locally connected.*

We only have to prove: If F is not locally connected, then F/A is not locally connected. For the local connectedness of C/A implies the local connectedness of $F/A = [C/A + S/A]$ and this will then imply the local connectedness of F .

So let us suppose that F is not locally connected. Then we can find a nucleus U of F , such that certain points of U arbitrarily near to 1 are not with 1 on a connected subset of U^2 . As S is locally connected U determines a connected nucleus V of S . As A is 0-dimensional V contains a subgroup A' of A , that is at the same time closed and open in A .† F/A' cannot be locally connected. This follows from: If two points a and b of U are separated in U^2 , then their images in F/A' are separated in the image of U . Suppose $U = U_a + U_b$, where U_a and U_b contain a and b and are separated in U^2 . Then their images are open and do not have a point in common, so they form a separation of the image of U between the images of a and b .

So if F is not locally connected then F/A' is not locally connected; but F/A' and F/A are locally isomorphic, so F/A is also not locally connected.

VII. The relation between S and S/A , F and F/A , D and $D/B = F$ is quite interesting. In order to have the simplest possible case we consider the relation between the direct product $P = [\bar{S}^{(1)} + \bar{S}^{(2)} + \dots]$ and $Q = [\bar{S}^{(1)} + \bar{S}^{(2)} + \dots]$. Then $Q = P/M$ where M is the 0-dimensional centrum of P . As direct product of connected, simply connected groups P is itself simply connected. We can make the fundamental group of Q into a topological group

* See second footnote on previous page.

† See E. R. van Kampen, *Annals of Mathematics*, vol. 36 (1935), no. 2, I, 4. The theorem goes back to L. Pontrjagin.

by combining into an arbitrary nucleus of the fundamental group all its elements isotopic with simple closed curves in an arbitrary nucleus of Q . It is then evident that the fundamental group of Q is the group M . Furthermore P can be defined as the universal covering group of Q . For any element in P corresponds to a class of isotopic arcs joining an element of Q to the identity element. A nucleus of P can now be determined as the collection of classes of arcs in Q isotopic with arcs in some nucleus U of Q .

These considerations indicate how a theory of covering spaces can be established for spaces in which arbitrarily small simple closed curves are not deformable into a point. This is quite independent of the fact that the spaces considered here are group spaces.

THE JOHNS HOPKINS UNIVERSITY.

THE INTERSECTION OF CHAINS ON A TOPOLOGICAL MANIFOLD.†

By WILLIAM W. FLEXNER.

1. In previous papers, one of them in collaboration with S. Lefschetz,‡ the author has dealt with topological manifolds. A topological manifold, M_n , is a compact separable Hausdorff space (therefore metric) which has a complete set of neighborhoods each of which is a combinatorial n -cell (F. M., p. 393). The following properties are shown in F. M. and F. M. 2 to hold for M_n : 1. the invariance of the homology characters; § 2. the standard properties of the Kronecker Index of two chains on M_n whose dimensions are p and $n - p$; 3. the Poincaré duality theorem. Property 1 was proved intrinsically, i. e. without imbedding M_n in a Euclidean space of higher dimension and using the properties of the space residual to M_n . In 2, however, the imbedding space was used to prove that every non-bounding p -cycle on M_n is cut by some $(n - p)$ -cycle on M_n with a Kronecker Index ± 1 . From 2 follows 3.

The present article makes no use of the imbedding theorem but defines intrinsically on M_n intersection cycles Γ_h ($h = p + q - n$), for two chains, C_p and C_q , on M_n of dimensionality p and q , not meeting one another's boundaries; and proves intrinsically that the cycles thus obtained form a locally homologous family (L. T., p. 183) about the geometric intersection, G , (L. T., p. 182) of C_p and C_q , thereby duplicating for M_n the salient theorem of the Lefschetz intersection theory for simplicial manifolds.

2. Some of the proofs to follow are complex. Therefore paragraphs 2-5 contain an outline describing without details the principal theorems and the methods used in their proof.

It is first shown that if M_n is orientable,¶ there is an orientable funda-

† Received December 15, 1934.

‡ S. Lefschetz and W. W. Flexner, *Proceedings of the National Academy of Sciences*, vol. 16 (1930), pp. 530-533; W. W. Flexner, *Annals of Mathematics*, (2), vol. 32 (1931), pp. 393-406 and pp. 539-548 (F. M., F. M. 2 in the sequel).

§ Terms and notation as in S. Lefschetz, "Colloquium lectures on topology," *American Mathematical Society Colloquium Publications*, vol. 12 (1930) (L. T. in the sequel).

¶ F. M., p. 399 *et seq.* On p. 400 the lines under the first formulas should read: "If the orientation of the cells E_n^i can be so chosen that for all i and j and all regions R , all the ε 's are positive, M_n is orientable with respect to the covering $\{E_n^i\}$, otherwise not."

mental cycle on M_n , and vice versa. If M_n is not orientable, the work is tacitly assumed to be carried out modulo 2. Then the construction of a typical intersection cycle, Γ_h , and the proof of the locally homologous family property are made. To define Γ_h , a covering of M_n by combinatorial n -cells, E_n^1, \dots, E_n^r , is now chosen and each E_n^i oriented concordantly with M_n . An intersection cycle, Γ_h , is then built up using this covering as follows. The chain C_p is deformed into a chain A_p^1 , the part of C_p not on E_n^1 being left invariant, the part on E_n^1 being deformed into a chain, C_p^1 , of the complex K_n^1 on E_n^1 . The deformation chain of the boundary is then added. Similarly C_q is deformed into A_q^1 except that the dual, K^{*n^1} , is used instead of K_n^1 . The part to be deformed is so chosen that its boundary is far from $F(E_n^1)$ ($F(A)$ means "boundary of A "). The chain C_q^{*1} is then defined as the subchain of A_q^1 on K^{*n^1} . As a result $F(C_p^1) \cdot C_q^{*1} = 0$. The chain $C_h^1 = C_p^1 \cdot C_q^{*1}$ is then defined as in L. T., ch. iv and it appears that $F(C_h^1) = C_p^1 \cdot F(C_q^{*1})$. Next the part of the intersection on E_n^2 is considered. The chain A_p^1 is deformed into A_p^2 just as C_p was into A_p^1 except that the deformation must be smaller, but A_q^1 is treated differently. Only the parts of A_q^1 on E_n^2 , far from $F(E_n^2)$ and not in C_q^{*1} are deformed onto K^{*n^2} ; the other points are left invariant. The deformation chain is added and C_q^{*2} is defined as the chain on K^{*n^2} . C_h^2 is then $C_p^2 \cdot C_q^{*2}$ and again $F(C_h^2) = C_p^2 \cdot F(C_q^{*2})$. By an inductive construction, this process is kept up until all cells covering the geometric intersection have been treated, thus giving fragmentary intersections: $C_h^1, C_h^2, \dots, C_h^r$.

3. It is now necessary to connect the boundaries of these fragments properly to make a cycle. If C_q^{*12} is the part of $F(C_q^{*1})$ in E_n^2 , then by the Lefschetz intersection theory, for every $\varepsilon > 0$, if the deformations producing A_p^2, A_q^2 are small enough, there is a singular chain, C_h^{12} , and a subchain, D_{q-1}^{*12} , of $F(C_q^{*2})$ such that $C_h^{12} \rightarrow C_p^1 \cdot C_q^{*12} - C_p^2 \cdot D_{q-1}^{*12} \pmod{M^{12}}$ near G where M^{12} is the ε -neighborhood of the complex carrying $C_p^1 \cdot F(C_q^{*12})$. A theorem of this type is then proved for an arbitrary pair of overlapping n -cells, E_n^i, E_n^j , $i < j$, so C_h^{ij} provides a connection $\pmod{M^{ij}}$ between parts of the boundaries of C_h^i and C_h^j . The cells of $F(C_h^i)$ not on the boundary of some C_h^{gi} or C_h^{ik} , $g < i < k$, can be shown to be in ΣM^{ab} , if the deformations are small enough. The neighborhood M^{ab} , defined analogously to M^{12} is an arbitrarily small one about the $(h-2)$ -complex carrying $C_p^a \cdot F(C_q^{*ab})$, which is determined at the a -th step of the construction. So $F(\sum_{i=1}^r C_h^i - \sum_{i,j=1}^r C_h^{ij}) = Q_{h-1}$ is an $(h-1)$ -chain in an arbitrarily small neighborhood of an $(h-2)$ -complex, and there

is a V_h such that $V_h \rightarrow Q_{h-1}$. Therefore $\Gamma_h = \Sigma C_h^i - \Sigma C_h^{ij} - V_h$ is a cycle and may be defined as an *intersection cycle* of C_p and C_q .

4. Clearly Γ_h , as defined, is a function of C_p, C_q , the n -cells E_n^i , and their order, and the sizes and characters of the various deformations. It can, however, be proved that any two intersection cycles derived from C_p and C_q are homologous in a preassigned neighborhood of G if the deformations giving rise to them are small enough, independent of the other factors.

To show this it is first proved that if Γ_h was obtained by small enough deformations, the chains giving rise to Γ_h can be further deformed to make an intersection cycle, Λ_h , on a covering $U_n, E_n^1, E_n^2, \dots, E_n^r$ where U_n is any n -cell of a covering of M_n . If the new deformations are small enough, $\Gamma_h \sim \Lambda_h$ close to G . This is the substance of Lemma 1 (No. 27) in the sequel. Repeated applications of Lemma 1 make it possible to derive from a given intersection another, homologous to it, on any other covering.

5. So it is sufficient for the general homology proof to show in addition to Lemma 1 that any two intersections, Γ_h and $\hat{\Gamma}_h$, on the same covering are homologous if they are obtained from C_p and C_q by small enough deformations. This is the substance of Lemma 2 (No. 27). The same notation is used as in the construction of Γ_h except that circumflex accents are used for quantities referring to $\hat{\Gamma}_h$. The proof will now be outlined. It is, roughly speaking,† the intersection of the final deforms A_p^r and A_q^r of C_p and C_q whose intersection gives Γ_h . Similarly an \hat{A}_p^r and an \hat{A}_q^r lead to $\hat{\Gamma}_h$. Because A_s^r and \hat{A}_s^r ($s = p, q$) originate from C_s there are chains $W_{s+1} \rightarrow A_s^r - \hat{A}_s^r$. The chain W_{p+1} can be deformed piece by piece onto K_n^1, K_n^2, \dots much as C_p was, leaving C_p^i and \hat{C}_p^i invariant for every i . Similarly W_{q+1} is deformed step by step onto $K_n^{*1}, K_n^{*2}, \dots$. If the part of W_{p+1} on K_n^k is C_{p+1}^k and that of W_{q+1} on K_n^{*k} is C_{q+1}^{*k} , then calculation of boundaries (L. T., p. 169) plus the fact that $F(C_p^k) \cdot C_q^{*k} = 0$ and $F(\hat{C}_p^k) \cdot \hat{C}_q^{*k} = 0$ gives that

$$(-1)^{n-q} C_{p+1}^k \cdot C_q^{*k} + \hat{C}_p^k \cdot C_{q+1}^{*k},$$

called C_{h+1}^k , is bounded by $C_p^k \cdot C_q^{*k} - \hat{C}_p^k \cdot \hat{C}_q^{*k} + X_h^k$. The chain X_h^k is a combination of chains near X where X is the corresponding combination reached at the preceding stages. This gives $C_{h+1}^k \rightarrow C_h^k - \hat{C}_h^k + X_h^k$. The simplicial parts of Γ_h and $\hat{\Gamma}_h$ are ΣC_h^k and $\Sigma \hat{C}_h^k$, so ΣC_{h+1}^k is bounded by these

† The statements that follow here are none of them exactly correct, but are made to bring out the general methods of the proof. The proof in detail is given in Nos. 31-34. In comparing the chains here with those of the same name in Nos. 31-34, it is, therefore, important to note that the correspondence is only schematic.

simplicial parts plus the X 's. A study of each X_h^k in relation to its predecessors similar to that of $C_p^2 \cdot F(D_{q-1}^{*12})$ in relation to $C_p^1 \cdot F(C_{q-1}^{*12})$ shows that the X 's and the non-simplicial parts of Γ_h and $\hat{\Gamma}_h$ can be used to make links between the pieces C_{h+1}^k in such a way as to give $\Gamma_h \sim \hat{\Gamma}_h$.

6. *Orientation of M_n .* Since M_n is connected, it follows, as in F. M. 2, p. 548, that there is one and only one independent non-bounding n -cycle, Γ_n , on M_n to a multiple of which every n -cycle is homologous. If M_n is orientable in the sense of F. M., p. 399, Γ_n will be oriented. Conversely, if Γ_n is an oriented cycle $\neq 0$ on M_n , then M_n is orientable according to F. M. with respect to any covering $E_n^1, E_n^2, \dots, E_n^r$. This is because the part of Γ_n on each E_n^i orients that E_n^i (see L. T., p. 44 and p. 101).

7. The next paragraphs deal with the definition of an intersection cycle for two chains. Being given two oriented chains C_p and C_q on M_n , assuming M_n orientable, such that $F(C_p)$ is nowhere nearer to C_q than $\alpha > 0$, and $F(C_q)$ is nowhere nearer to C_p than α , it is desired to find a semi-simplicial cycle, Γ_h , (F. M., p. 540) of dimensionality $h = p + q - n$ arbitrarily near the geometric intersection, G , and playing the rôle of an "intersection cycle."

8. Let $E_n^1, E_n^2, \dots, E_n^r$ be the subset covering G of a covering of M_n (F. M., p. 395). There will, by definition of a covering, be a $\beta > 0$ such that every point of G has on M_n a neighborhood around it in some E_n^i of the subset with diameter β .

9. *Fundamental construction.* Choose a $\delta > 0$. Step 1. On E_n^1 take a complex K_n^1 of mesh (L. T., p. 85) $\varepsilon_1/2 < \alpha/20r$ and $< \beta/20r$, where r is as defined in No. 8. If K_n^{*1} is the dual (L. T., p. 132) on E_n^1 of K_n^1 , it is of mesh ε_1 . Subdivide the chains C_p and C_q until the mesh of their cells is $\varepsilon_1/2$ and call the subdivided chains by the same names again. Next deform C_p into a chain A_p^1 by means of an $\varepsilon_1/2$ -deformation, as follows. Leave unaltered the closed p -cells of C_p not entirely on K_n^1 . Deform the remainder onto a sub-chain of K_n^1 and call the new chain on K_n^1 , C_p^1 . Add the deformation chain of the boundary of the piece which was deformed.

10. Deform C_q in the same manner, using K_n^{*1} instead of K_n^1 , and leaving invariant all q -cells of C_q not on E_n^1 and not at a distance of more than $4\varepsilon_1$ from $F(E_n^1)$. Add the deformation chain; call the deformed chain A_q^1 , and the part of A_q^1 on K_n^{*1} , C_q^1 . Let $C_h^1 = C_p^1 \cdot C_q^1$, $h = p + q - n$. If ε_1 is small enough, C_h^1 is within δ of G .

11. Since all points of $F(C_p^1)$ must lie within ε_1 of $F(E_n^1)$, or by choice of ε_1 (see No. 7) be far from C_q , no point of C_q^{*1} can meet $F(C_p^1)$. Therefore (L. T., p. 169):

THEOREM A^1 . $F(C_k^1) = C_p^1 \cdot F(C_q^{*1})$.

12. Assume steps $2, 3, \dots, k-1$ to have been taken, Theorems A^2, \dots, A^{k-1} to have been proved, and, for $i < j < k$, the following chains to have been defined: $A_p^j, A_q^j, sA_p^j, sA_q^j, C_p^j, C_q^{*j}, C_k^j, C_{q-1}^{*ij}, D_{q-1}^{*ij}, R_q^j, \Delta_q^{ij}$. Let C_{q-1}^{*ik} be the chain sum of the closed $(q-1)$ -cells of $F(C_q^{*i})$, $i < k$, which are entirely in E_n^k with no point within $4\varepsilon_1$ of $F(E_n^k)$, and which have no interior point in C_{q-1}^{*ij} or D_{q-1}^{*aj} , $a < i < j < k$.†

13. Step k . Take on E_n^k a complex, K_n^k , of mesh ε_k , where $6\varepsilon_k < \varepsilon_{k-1}$, and ε_k satisfies other conditions to be specified later (Nos. 15, 16 and Theorems B, C, D). Let K_n^{*k} be the dual of K_n^k . Subdivide the chains A_a^{k-1} , $a = p, q$, into chains, sA_a^{k-1} of mesh $\varepsilon_k/2$.

14. Now deform sA_p^{k-1} into A_p^k just as, in No. 10, C_p was deformed into A_p^1 , using K_n^k instead of K_n^1 , and call the part of the new chain on K_n^k , C_p^k .

15. By an ε_k -deformation carry sA_q^{k-1} into a chain A_q^k : the deformation to be as follows. It shall carry a chain R_q^{k-1} into a subcomplex, C_q^{*k} , of K_n^{*k} . The chain R_q^{k-1} is made up of the closed q -cells of sA_q^{k-1} in E_n^k and at a distance of more than $4\varepsilon_1$ from $F(E_n^k)$, but minus the cells which are 1) in C_q^{*i} , $i < k$; 2) in Δ_q^{ij} , the deformation chain joining C_q^{*ij} and D_{q-1}^{*ij} , $i < j < k$; plus 3) such cells of sA_q^{k-1} in E_n^k and not in 1) or 2) as have, for some j , a point of the subdivided C_q^{*jk} but no $(q-1)$ -cell of the subdivided C_{q-1}^{*js} or D_{q-1}^{*sj} , $s < k$, on their boundaries. In other words, R_q^{k-1} is the chain sum of the closed q -cells of sA_q^{k-1} well inside E_n^k , with C_{q-1}^{*jk} on its boundary for every j , but no $(q-1)$ -cells of C_{q-1}^{*js} or D_{q-1}^{*sj} on its boundary.

All points not in R_q^{k-1} are left invariant. Add the deformation chain of the boundary of R_q^{k-1} . Let $C_k^k = C_p^k \cdot C_q^{*k}$. Assuming that at each previous stage C_k^i was within $i\delta$ of G , $i < k$, by taking ε_k small enough, C_k^k may be brought within $k\delta$ of G , justifying the assumption.

16. Let D_{q-1}^{*ik} , $i < k$, be the image in $F(C_q^{*k})$, under the deformation just defined, of C_{q-1}^{*ik} . By condition 3, No. 15, this image exists. Take ε_k so small that no cell of D_{q-1}^{*ij} is within $2\varepsilon_1$ of $F(E_n^k)$.

† As defined, C_{q-1}^{*ik} is the part of the boundary of $F(C_q^{*i})$ which is in E_n^k but not in E_n^j , $j < k$.

17. All points of $F(C_p^k)$ must be within ε_k of $F(E_n^k)$ or else far from C_q^{*k} , for the deformations are too small to bring images of $F(C_p)$ and C_q together. So, since C_q^{*k} is entirely farther than $2\varepsilon_1 > \varepsilon_k$ from $F(E_n^k)$, $F(C_p^k) \cdot C_q^{*k} = 0$. Therefore (L. T., p. 169):

THEOREM A^k . $F(C_n^k) = C_p^k \cdot F(C_q^{*k})$.

The construction and proof given here is carried out until, at the r -th stage, all n -cells $E_n^1, E_n^2, \dots, E_n^r$ have been treated.

18. THEOREM B . If $1 \leq i < j \leq r$ and if $\varepsilon_{i+1}, \varepsilon_{i+2}, \dots, \varepsilon_j$ are small enough, then $C_p^i \cdot C_{q-1}^{*ij} \sim C_p^j \cdot D_{q-1}^{*ij} \bmod M^{ij}$ on N^{ij} , where M^{ij} is a τ^{ij} -neighborhood of $|C_p^i \cdot F(C_{q-1}^{*ij})|$, τ^{ij} arbitrary (but is to be chosen $< \delta$), and N^{ij} is a ρ^i -neighborhood of $|C_p^i \cdot C_{q-1}^{*ij}|$. The values of τ^{ij} give a maximum value to ρ^i , but ρ^i approaches zero with $\varepsilon_{i+1}, \varepsilon_{i+2}, \dots, \varepsilon_j$ independent of τ^{ij} , and so can be taken $< \delta$.

19. Note that $\varepsilon_{i+1}, \varepsilon_{i+2}, \dots, \varepsilon_j$ are determined after the i -th step of the fundamental construction (referred to in the sequel as f. c.) whereas $|C_p^i \cdot F(C_{q-1}^{*ij})|$ was determined and fixed previously, at the i -th step.

20. Proof of B . Both $C_p^i \cdot C_{q-1}^{*ij}$ and $C_p^j \cdot D_{q-1}^{*ij}$ are intersections in the sense of L. T., ch. iv, of the chains C_p^i and C_{q-1}^{*ij} which do not meet one another's boundaries modulo M^{ij} . They are, therefore, homologous as stated if the ε 's are small enough. Since the distance from $A_p^i - C_p^i$ to C_{q-1}^{*ij} is greater than zero and depends on the ε 's, no points of A_p^i not in C_p^i can have images in A_p^j meeting D_{q-1}^{*ij} provided that the ε 's are small enough.

21. Now let E_{h-1} be a closed $(h-1)$ -cell of $F(C_n^k)$, $1 \leq k \leq r$, and suppose $E_{h-1} = C_p^k \cdot E_{q-1}^{*k}$ where E_{q-1}^{*k} is a closed $(q-1)$ -cell of $F(C_q^{*k})$. Further call E_{q-1} any one of the closed $(q-1)$ -cells of sA_q^{k-1} of which E_{q-1}^{*k} is an image. There are point sets, \mathcal{E} , of which E_{q-1} is image in each A_q^a , $1 \leq a < k-1$, and because regular subdivision was used in f. c., no \mathcal{E} has points in more than one closed q -cell of A_q^a . Let E_q^a be a closed q -cell of A_q^a carrying an \mathcal{E} .

22. THEOREM C . If $1 \leq k \leq r$, all cells E_{h-1} which are not cells of $C_p^k \cdot D_{q-1}^{*k}$ or in M^{ij} , $i < j < k$, are cells of some $C_p^k \cdot C_{q-1}^{*ks}$, $k < s \leq r$, provided $\varepsilon_{i+1}, \varepsilon_{i+2}, \dots, \varepsilon_k$ are small enough.

† Following the recent usage of S. Lefschetz: if A is a simplicial chain, $|A|$ is the complex carrying A .

Proof. In order to show that E_{h-1} is in some $C_p^k \cdot C_{q-1}^{*ks}$ it is, because of f. c., sufficient to show that E_{q-1} is not in an $F(C_q^j)$, $j < k$, and that E_{h-1} is within some E_n^s by a distance of at least $4\varepsilon_1$.

Because of the condition of No. 7 and the smallness of $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k$, neither $F(C_p)$ nor $F(C_q)$ nor their deforms play any rôle in $F(C_h^k)$. So for E_{h-1} to be in $F(C_h^k)$, E_{q-1} must either be

- 1) a cell in $F(C_q^j)$, $j < k$, (now denoting chain and subdivision by $F(C_q^j)$);
- 2) a cell of $F(\Delta_q^{ij})$, $i < j < k$, (see No. 15) or the image of such a cell;
- or 3) a cell within $4\varepsilon_1 + \varepsilon_k$ of $F(E_n^k)$ and not in 1) or 2).

In case 1), every cell of $F(C_q^j)$ either belongs to C_{q-1}^{*jk} or is not deformed onto K_n^{*k} (condition 3, No. 15). Therefore the images of such cells are in D_{q-1}^{*jk} .

In case 2) let $\Delta_q'^{ij}$ be the image of Δ_q^{ij} and let $\Delta_{q-1}^{ij} = F(\Delta_q'^{ij}) - (C_{q-1}^{*ij} - D_{q-1}^{*ij})$. Now let $\Delta_{q-1}^{*ij,k}$ be the part of Δ_{q-1}^{ij} in $F(C_q^k)$. If $C_q^k \cdot \Delta_{q-1}^{*ij,k}$ did not lie in M^{ij} when the ε 's are small enough, there would be for each of an infinite number of sets of these ε 's as they approached zero, a point, P , of the corresponding $C_p^k \cdot \Delta_{q-1}^{*ij,k}$ at a distance, $d(P) > \rho > 0$ from $|C_p^i \cdot F(C_{q-1}^{*ij})|$, which complex is not a function of the ε 's mentioned. The points P would then have a limit point, L , at a distance $\geq \rho$ from $|C_p^i \cdot F(C_{q-1}^{*ij})|$.

i. Suppose L is not on $|F(C_{q-1}^{*ij})|$ but has a distance $d' > 0$ from it. The chain Δ_{q-1}^{ij} being the image of the deformation chain of $F(C_{q-1}^{*ij})$ would, if $\varepsilon_j + \varepsilon_{j+1} + \dots + \varepsilon_k < d'/4$, lie within $d'/4$ of $|F(C_{q-1}^{*ij})|$, so $C_p^k \cdot \Delta_{q-1}^{*ij,k}$, a subset, would also; and the points P would be, after a certain one, all within $d'/2$ of $|F(C_{q-1}^{*ij})|$ contrary to the hypothesis that L is their limit point. Thus case i. cannot occur.

ii. Suppose, then, that L is on $|F(C_{q-1}^{*ij})|$. Since by f. c., no points of A_p^i not on C_p^i can meet $|F(C_{q-1}^{*ij})|$, it is possible by taking the ε 's small enough to bring the point set intersection of $|C_p^k|$ and $|F(C_{q-1}^{*ij})|$ arbitrarily close to $|C_p^i \cdot F(C_{q-1}^{*ij})|$. Then the points P , since each is on a C_p^k and near $|F(C_{q-1}^{*ij})|$, will again, after a certain one, be nearer by a finite amount to $|C_p^i \cdot F(C_{q-1}^{*ij})|$ than L is. So since i. and ii., are exhaustive, $C_p^k \cdot \Delta_{q-1}^{*ij,k}$ must be in M^{ij} .

In case 3) there must be a first n -cell, E_n^a , $a \neq k$ containing E_{q-1} in such a way that all its points are at a distance of at least $\beta/2$ from $F(E_n^a)$ (condition of No. 8). If $a > k$ the theorem is proved. If $a < k$ there is in E_n^a and E_q^a of the type defined in No. 21 which must be inside E_n^a by a margin of $[\beta/2 - (\varepsilon_k + \varepsilon_{k-1} + \dots + \varepsilon_a)] > 4\varepsilon_1$. Therefore E_q^a is a cell of $F(C_q^j)$, $j \leq a$, or Δ_q^{xy} , $x < y < a$. This reduces case 3) to cases 1) and 2) already considered, since the only cells of Δ_q^{xy} deformed are on $F(\Delta_q^{xy})$.

23. THEOREM D. If E_{h-1} is in $C_p^k \cdot D_{q-1}^{*ik}$ and in $C_p^k \cdot D_{q-1}^{*jk}$, $i < j < k$, it is in M^{ij} provided $\varepsilon_{i+1}, \varepsilon_{i+2}, \dots, \varepsilon_k$ are small enough.

If E_{q-1}^i and E_{q-1}^j are the originals of E_{q-1}^{*ik} in C_{q-1}^{*ik} and C_{q-1}^{*jk} respectively, then they must be within ε_k of each other. Then E_{q-1}^j must be within $2\varepsilon_j$ of $|F(C_{q-1}^{*ik})|$; for, since ε_k is less than the meshes of both K_n^{*i} and K_n^{*j} , it is only by being in a q -cell of C_{q-1}^{*jk} abutting on $|F(C_{q-1}^{*ik})|$ that E_{q-1}^j can be within ε_k of C_{q-1}^{*ik} . Therefore E_{q-1}^j is within $2\varepsilon_j + \varepsilon_k$ of $|F(C_{q-1}^{*ik})|$. But E_{h-1} is also on a part of C_p^k which was obtained by an $\varepsilon_{i+1} + \varepsilon_{i+2} + \dots + \varepsilon_k$ deformation from C_p^i . Therefore if these ε 's are small enough, E_{q-1} is within τ^{ik} of $|C_p^i \cdot F(C_{q-1}^{*ik})|$, i. e. in M^{ik} . (Note once more that M^{ik} is independent of $\varepsilon_{i+1}, \varepsilon_{i+2}, \dots, \varepsilon_k$).

24. THEOREM E.

$$F(C_h^k) = - \sum_{i=1}^{k-1} C_p^k \cdot D_{q-1}^{*ik} + \sum_{i=k+1}^r C_p^k \cdot C_{q-1}^{*ki} \bmod \Sigma M^{ab}.$$

It is a consequence of f. c. and Theorems C and D that each cell of $F(C_h^k)$ not in M^{ab} is in one and only one of the chains on the right-hand side of the formula above. It remains to make sure of the coefficients in each case. Those cells in the second term have the right coefficient by Theorem A^{*} and the definition of C_{q-1}^{*ki} . As to the first term, since C_q is an oriented chain, D_{q-1}^{*ik} is negatively related to $F(C_q^{*k})$, so $C_p^k \cdot D_{q-1}^{*ik}$ is negatively related to $F(C_h^k)$.

25. The condition of No. 8 makes it sure that f. c. comes to an end at the r -th step: all $(h-1)$ -cells of $C_p^r \cdot F(C_q^{*r})$ belong either in $-C_p^r \cdot D_{q-1}^{*jr}$, $j < r$, or in $M = \Sigma M^{ab}$ since there can be no C_{q-1}^{*rs} , $s > r$.

Form the sum

$$C_h = \sum_{k=1}^r C_h^k - \sum_{i,k=1}^{r(i < k)} C_h^{ik}$$

where $C_h^{ik} \rightarrow C_p^i \cdot C_{q-1}^{*ik} - C_p^k \cdot D_{q-1}^{*ik} \bmod M^{ik}$ on N^{ik} .

The existence of C_h^{ik} follows from Theorem B. Since it is within $(r+1)\delta$ of G , C_h is arbitrarily close to G . Computing the boundary of C_h formally (L. T., p. 169) and using Theorems B and E and the Theorems A, gives that $F(C_h)$ is an $(h-1)$ -cycle on M . But M is an arbitrarily small neighborhood of an $(h-2)$ -complex, so, as in F. M. 2, p. 541, $F(C_h) \sim 0$ on M' where M' is a neighborhood of M whose size approaches zero with the size of M , and whose distance from G approaches zero with the size of M . Therefore there is a complex, V_h , on M' such that $V_h \rightarrow F(C_h)$.

26. Then $\Gamma_h = C_h - V_h$ is a semi-simplicial h -cycle on M_n arbitrarily close to G . The cycle is defined as an *intersection cycle on the covering* $E_n^1, E_n^2, \dots, E_n^r$ of the chains C_p and C_q .

27. The next numbers will be devoted to the proof of the following theorem.

THEOREM F. *Two intersection cycles Γ_h and $\hat{\Gamma}_h$ of the chains C_p and C_q are homologous in any arbitrarily small given neighborhood of G provided the deformations used in getting them are small enough, even if Γ_h is on the covering $E_n^1, E_n^2, \dots, E_n^r$, and $\hat{\Gamma}_h$ is on a different covering, $H_n^1, H_n^2, \dots, H_n^s$.*

If M_n is orientable, E_n^i and H_n^j must be oriented concordantly with the fundamental cycle on M_n . Otherwise the work is done modulo 2. It should be noted that Γ_h and $\hat{\Gamma}_h$ are said to be *on the same covering* if the n -cells and their order are the same, and if on each E_n^i the same fundamental complex K_n^i , is used in obtaining Γ_h and $\hat{\Gamma}_h$. Otherwise the coverings are termed different. The proof of Theorem F depends on two lemmas.

LEMMA 1. *If Γ_h is an intersection cycle on the covering $E_n^1, E_n^2, \dots, E_n^r$ and obtained by small enough deformations, and U_n is another n -cell of some covering of M_n , and $\varepsilon > 0$ is an arbitrary number; then there exists an intersection cycle, Λ_h , on the covering $U_n, E_n^1, E_n^2, \dots, E_n^r$, and such that $\Gamma_h \sim \Lambda_h$ within ε of G .*

LEMMA 2. *If Γ_h and $\hat{\Gamma}_h$ are both on the covering $E_n^1, E_n^2, \dots, E_n^r$, and $\varepsilon > 0$ is given; then, if the deformations producing Γ_h and $\hat{\Gamma}_h$ are small enough, $\Gamma_h \sim \hat{\Gamma}_h$ within ε of G .*

28. It will now be shown that Theorem F follows from the lemmas. If Γ_h is on $E_n^1, E_n^2, \dots, E_n^r$, and $\hat{\Gamma}_h$ is on $H_n^s, E_n^1, E_n^2, \dots, E_n^r$; then $\Gamma_h \sim \hat{\Gamma}_h$ with proper stipulations as to size of deformations etc. By Lemma 1 there is a cycle, Λ_h , on the second covering such that $\Gamma_h \sim \Lambda_h$. Then, again with proper stipulations, Lemma 2 gives $\Lambda_h \sim \hat{\Gamma}_h$; from which follows $\Gamma_h \sim \hat{\Gamma}_h$.

29. If $\hat{\Gamma}_h$ is on $H_n^1, H_n^2, \dots, H_n^s, E_n^1, E_n^2, \dots, E_n^r$ and Γ_h is on $E_n^1, E_n^2, \dots, E_n^r$; then $\Gamma_h \sim \hat{\Gamma}_h$. This result is obtained by repeated use of the argument of No. 28. But since $H_n^1, H_n^2, \dots, H_n^s$ covers G , if the deformations are small enough the compound covering is equivalent to $H_n^1, H_n^2, \dots, H_n^s$ from the point of view of intersections. The statement at the head of this number is thus equivalent to Theorem F.

30. *Proof of Lemma 1.* The chains to be deformed to get Λ_h from Γ_h are A_p^r and A_q^r . These, it should be recalled, are the final deforms of C_p and C_q used in getting Γ_h . Starting with these chains begin, on U_n , to build up Λ_h in the same way that Γ_h was built up from C_p and C_q in f. c. The simplicial piece of Λ_h on U_n will be in part deforms of parts of the A 's which gave rise to simplicial pieces of Γ_h . So if the additional deformations are small enough (from Γ_h to Λ_h) and Γ_h itself was got by small enough deformations, L. T., ch. iv shows there are homologies within $\varepsilon/2$ of G between corresponding parts of Γ_h and Λ_h mod neighborhoods, N , of $(h-1)$ -complexes on Γ_h of the type $|C_p^i \cdot C_{q-1}^{*ij}|$ (see No. 10). These neighborhoods depend in size on the parts of Γ_h on cells, E_n^k , $i < k$, and on the additional deformations used to get Λ_h , so they can be arbitrarily small.

If Γ_h^u is the sum of the closed h -cells of Γ_h on U_n , and N' is a suitable neighborhood of the complex carrying the simplicial part of $F(\Gamma_h^u)$; then $\Gamma_h \sim \Lambda_h$ mod $N + N' + (M_n - U_n)$; and the diameter of N' approaches zero as the deformations producing Λ_h from Γ_h approach zero. Outside U_n , Λ_h can be identical with Γ_h so $\Gamma_h \sim \Lambda_h$ mod $(N + N')$. Since $N + N'$ is an arbitrarily small neighborhood of an $(h-1)$ -complex, $\Gamma_h \sim \Lambda_h$ as stated in Lemma 1 (see F. M. 2, p. 541).

31. *Proof of Lemma 2* (see No. 5). The proof involves the construction of an $(h+1)$ -chain, C_{h+1} on M_n such that $C_{h+1} \rightarrow \Gamma_h - \hat{\Gamma}_h$ within ε of G . This construction is similar to f. c. and is to be made by induction. In what follows the notation of f. c. will be used for Γ_h . The same notation with a circumflex accent ($\hat{}$) added will be used when $\hat{\Gamma}_h$ is in question.

Step 1. The actual proof proceeds as follows. Choose an $\eta > 0$. Since A_p^1 and \hat{A}_p^1 are both deforms of C_p , there is a chain, $W_{p+1}^1 \rightarrow A_p^1 - \hat{A}_p^1$ on M_n of mesh $\varepsilon_1/2$. Similarly there is a $(q+1)$ -chain, $W_{q+1}^1 \rightarrow A_q^1 - \hat{A}_q^1$. By an $\varepsilon_1/2$ -deformation carry the closed $(p+1)$ -cells of W_{p+1}^1 entirely on K_n^1 into a subchain, C_{p+1}^1 , of K_n^1 , leaving A_p^1 and \hat{A}_p^1 invariant. This is possible because the requisite parts of the p -chains are already on K_n^1 . Add the deformation chain of the boundary of the piece which was deformed, and call the entire new $(p+1)$ -chain A_{p+1}^1 .

Deform W_{q+1}^1 in the same manner using K_n^{*1} instead of K_n^1 and leaving invariant all q -cells of W_{q+1}^1 not in E_n^1 and not at a distance of more than $4\varepsilon_1$ from $F(E_n^1)$, unless they have a point of C_q^{*1} or \hat{C}_q^{*1} on their boundaries in which case they are deformed with the rest. The chains \hat{C}_q^{*1} and C_q^{*1} themselves are to be left invariant. Add the deformation chain of the boundary and call the total deformed chain A_{q+1}^1 and the part of A_{q+1}^1 on K_n^{*1} , C_{q+1}^{*1} . The chain $X_q^{*1} = F(C_{q+1}^{*1}) - (C_q^{*1} - \hat{C}_q^{*1})$ is then nowhere

nearer than $2\varepsilon_1$ to $F(E_n^1)$, whereas $X_p^1 = F(C_{p+1}^1) - (C_p^1 - \hat{C}_p^1)$ is nowhere farther than ε_1 from $F(E_n^1)$.

$$\text{Let } C_{h+1}^1 = (-1)^{n-q} C_{p+1}^1 \cdot C_q^{*1} + \hat{C}_p^1 \cdot C_{q+1}^{*1},$$

intersections being of chains on K_n^1 and K_n^{*1} , are here meant in the sense of L. T., ch. iv, § 1. Then if ε_1 is small enough L. T., pp. 169 and 187 give

$$C_{h+1}^1 \rightarrow C_h^1 - \hat{C}_h^1 + X_h^1, \text{ where, since } F(\hat{C}_p^1) \cdot C_{q+1}^{*1} = 0,$$

$$X_h^1 = (-1)^{n-1} C_{p+1}^1 \cdot F(C_q^{*1}) + \hat{C}_p^1 \cdot X_q^{*1}.$$

32. Assume steps 2, 3, ..., $k-1$ to have been made and the necessary chains to have been defined for i and j , $i < j < k$. Let X_q^{*jk} be the chain sum of the closed q -cells of X_q^{*j} which are 1) entirely on E_n^k , and 2) have no point within $4\varepsilon_1$ of $F(E_n^k)$ unless a $(q-1)$ -cell of C_q^{*jk} or \hat{C}_{q-1}^{*jk} is on their boundary, and 3) have no interior points on X_q^{*jg} or Y_q^{*ij} , $g < k$. The chain X_q^{*jk} plays the rôle in this proof of C_q^{*jk} in f. c. It is, roughly speaking, the part of X_q^{*j} in E_n^k but not in E_n^j , and is so defined that all cells of C_q^{*jk} and \hat{C}_{q-1}^{*jk} are on its boundary.

Step k is then made as follows. As a consequence of step $k-1$ of the induction and step k of f. c., there are on M_n chains W_{p+1}^{k-1} and W_{q+1}^{k-1} such that $W_{s+1}^{k-1} \rightarrow A_s^k - \hat{A}_s^k$ ($s = p, q$). Subdivide the cells of W_{p+1}^{k-1} and W_{q+1}^{k-1} until their mesh is $\varepsilon_k/2$ and ε_k respectively without altering the A 's which already satisfy this condition. Call the new chains sW_{p+1}^{k-1} and sW_{q+1}^{k-1} .

Now carry sW_{p+1}^{k-1} into A_{p+1}^k just as W_{p+1}^1 was carried into A_{p+1}^1 . Call the new $(p+1)$ -chain on K_n^k C_{p+1}^k .

Next sW_{q+1}^{k-1} is to be treated.

LEMMA. If sW_{q+1}^{k-1} contains a $(q+1)$ -cell, E_{q+1} having a q -face, E_q , in X_q^{*js} or Y_q^{*sj} , $s < j < k$, and a q -face E'_q in R_q^{k-1} (see No. 15), a resubdivision of sW_{q+1}^{k-1} will avoid this without creating new situations of the same type.

Proof. If E_q and E'_q have a $(q-1)$ -face in common, it is, by construction, a $(q-1)$ -cell of C_q^{*js} or D_q^{*sj} . But by definition of R_q^{k-1} this is impossible. So a subdivision of E_{q+1} by section (L. T., p. 68) will avoid the situation in the desired way. A similar lemma holds for \hat{R}_q^{k-1} .

This lemma shows that if the ε_k -subdivisions are small enough the following definition of R_{q+1}^{k-1} is self-consistent. The chain R_{q+1}^{k-1} is the set of all closed q -cells of sW_{q+1}^{k-1} in E_n^k and at a distance of more than $4\varepsilon_1$ from $F(E_n^k)$ minus the cells which are

1) in C_{q+1}^{*j} , $j < k$;
 2) in Δ_{q+1}^{ij} , the deformation chain joining X_{q+1}^{*ij} and Y_{q+1}^{*ij} ;
 plus 3) such q -cells of sW_{q+1}^{k-1} in E_n^k but not in 1) or 2) as have on their boundaries,

a) for some j a point of X_{q+1}^{*jk} but no $(q-1)$ -cell of X_{q+1}^{*js} or Y_{q+1}^{*js} , $s < k$, (see No. 15),

or b) a q -cell of R_q^{k-1} or \hat{R}_q^{k-1} .

By an ε_k deformation carry R_{q+1}^{k-1} into a sub-chain, C_{q+1}^{*k} , of K_n^{*k} . Add the deformation chain of the boundary of the deformed part and call the new chain A_{q+1}^{*k} . Let Y_{q+1}^{*ik} , $i < k$, be the image under this deformation of X_{q+1}^{*ik} . Take ε_k so small that no cell of Y_{q+1}^{*ik} is within $2\varepsilon_i$ of $F(E_n^k)$. This new chain plays here approximately the rôle of D_{q-1}^{*ik} in f. c. Now the chain $X_{q+1}^{*k} = F(C_{q+1}^{*k}) - (C_{q+1}^{*k} - \hat{C}_{q+1}^{*k})$ is nowhere nearer than ε_1 to $F(E_n^k)$, whereas $X_p^k = F(C_{p+1}^k) - (C_{p+1}^k - \hat{C}_p^k)$ is nowhere farther than ε_k from $F(E_n^k)$.

33. Let $C_{h+1}^k = (-1)^{n-q} C_{p+1}^k \cdot C_{q+1}^{*k} + \hat{C}_p^k \cdot C_{q+1}^{*k}$. Then if ε_k is small enough, $C_{h+1}^k \rightarrow C_h^k - \hat{C}_h^k + X_h^k$ within $k\eta$ of G , where, since $F(\hat{C}_p^k) \cdot C_{q+1}^{*k} = 0$, $X_h^k = (-1)^{n-q} C_{p+1}^k \cdot F(C_{q+1}^{*k}) + \hat{C}_p^k \cdot X_{q+1}^{*k}$. Define X_h^{ik} , Y_h^{ik} , L_h^{ik} , $i < k$, as follows:

$$\begin{aligned} X_h^{ik} &= (-1)^{n-q} C_{p+1}^i \cdot C_{q-1}^{*ik} + \hat{C}_p^i \cdot X_{q-1}^{*ik}; \\ Y_h^{ik} &= (-1)^{n-q} C_{p+1}^k \cdot D_{q-1}^{*ik} + \hat{C}_p^k \cdot Y_{q-1}^{*ik}; \end{aligned}$$

L^{ik} is the λ_i -neighborhood ($0 < \lambda_i < \eta$) of

$$|\hat{C}_p^i \cdot F(X_{q+1}^{*ik})| + |\hat{C}_p^i \cdot C_{q-1}^{*ik}| + |C_p^i \cdot C_{q-1}^{*ik}| + |\hat{C}_p^i \cdot \hat{C}_{q-1}^{*ik}|.$$

Take λ_i so large that L^{ik} includes the neighborhoods N^{ik} , M^{ik} , \hat{N}^{ik} and \hat{M}^{ik} . The diameter λ_i approaches zero as δ does, (see f. c.). Note that L^{ik} does not depend on ε_{i+1} , ε_{i+2} , \dots , ε_k .

THEOREM G. If $1 \leq k \leq r$, for every λ_i , the quantities ε_{i+1} , ε_{i+2} , \dots , ε_k can be taken so small that there exists an $(h+1)$ -chain, C_{h+1}^{ik} , within $(k+1)\eta$ of G such that

$$C_{h+1}^{ik} \rightarrow X_h^{ik} - Y_h^{ik} \bmod L^{ik}.$$

Proof. The intersections $C_{p+1}^i \cdot C_{q-1}^{*ik}$ and $C_{p+1}^k \cdot D_{q-1}^{*ik}$ are in the sense of L. T., ch. iv, considering as the original chains C_{p+1}^i and C_{q-1}^{*ik} which do not meet one another's boundaries mod L^{ik} , so the proof goes through like Theorem B, No. 18, as far as this pair is concerned. The same sort of proof holds for \hat{C}_p^i and X_{q+1}^{*ik} .

34. The construction of the preceeding paragraphs has been so made that all h -cells of X_h^k are either images of h -cells in X_h^i , $i < k$, or within $5\varepsilon_1$ of $F(E_n^k)$. The construction is also such that Theorems C and D (Nos. 22, 23) hold for $C_{p+1}^k \cdot F(C_q^{*k})$ and $\hat{C}_p^k \cdot X_q^{*k}$ and the neighborhoods L^{ij} as they do for $C_p^k \cdot F(C_q^{*k})$ and the neighborhoods M^{ij} . This is because the construction of C_{p+1}^k and X_q^{*k} is exactly analogous to that of C_p^k and $F(C_q^{*k})$. Combining these results for $C_{p+1}^k \cdot F(C_q^{*k})$ and $\hat{C}_p^k \cdot X_q^{*k}$ gives:

THEOREM C' . *If $1 \leq k \leq r$, all h -cells of X_h^k which are not cells of Y_q^{*ik} are in L^{ij} , $i < j < k$, or in some X_q^{*ks} $k < s \leq r$, provided the ε 's are small enough.*

THEOREM D' . *If an h -cell of X_h^k is in Y_q^{*ik} and Y_q^{*jk} , $i < j < k$, it is in L^{ik} provided the ε 's are small enough.*

As before, L^{ij} is always fixed after $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_i$ have been determined.

These two theorems combine to give the analogue, E' , of Theorem E (No. 24). Theorem E' plus the fact that the process here considered comes to an end at the r -th step, gives $C'_{h+1} \rightarrow (C_h^i - \hat{C}_h^i) \bmod \Sigma L^{ab}$ within $(r+1)\eta$ of G , where $C'_{h+1} = \Sigma C_{h+1}^i - \Sigma C_{h+1}^{ab}$. Since $L^{ab} \supset N^{ab}$ and \hat{N}^{ab} , $C'_{h+1} \rightarrow \Gamma_h - \hat{\Gamma}_h + Q_h$, where Q_h is an h -cycle on ΣL^{ab} . If $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_r$ are small enough, $Q_h \sim 0$ within $(r+2)\eta$ of G , so there is an $(h+1)$ -chain

$$C_{h+1} \rightarrow \Gamma_h - \hat{\Gamma}_h$$

within $(r+2)\eta$ of G . If $\eta = \varepsilon/(r+2)$ this proves Lemma 2, and completes the program outlined in Nos. 1-5.

ON THE IMBEDDING OF METRIC SETS IN EUCLIDEAN SPACE.*

By W. A. WILSON.

1. It is the purpose of this note to make a slight extension of results previously obtained by the writer † and to give a modification of Menger's general conditions for the imbedding of n points of a metric space in Euclidean space.

With regard to the first topic it is proved on pp. 515-16 of the paper mentioned that a complete space, which is convex and externally convex and has the four-point property, has the n -point property for every integer n . We now proceed to show that the requirement of external convexity is needless.

Using the notation of this proof, let $T_1 \simeq T'_1$, $T_0 \simeq T'_0$, and $T_{01} \simeq T'_{01}$. If the line through a'_0 and a'_1 meets T'_{01} , external convexity is not needed for the proof as given.‡ In the opposite case it is clear that there is a point u' in T'_1 near enough to the centroid of T'_{01} so that: if a'_0 and a'_1 are on the same side of E_{n-2} and $a'_0 u'$ is produced to meet T'_{01} in x' , $u' x'$ lies also in T'_0 ; and, if a'_0 and a'_1 are on opposite sides of E_{n-2} , $a'_1 u'$ cuts T'_{01} . In the congruence $T_1 \simeq T'_1$, let $u \sim u'$. Then by the argument of p. 516 the points u, a_1, a_2, \dots, a_n can be imbedded in E_n and the sign of the determinant $D(u, a_1, a_2, \dots, a_n) \neq \text{sign}(-1)^n$.

Let us now suppose that $\text{sign } D(a_0, a_1, \dots, a_n) = \text{sign}(-1)^n$. Since D is a continuous function of each variable and changes sign when a_0 is replaced by u , there is some point v on the segment ua_0 for which $D(v, a_1, a_2, \dots, a_n) = 0$. Then the points v, a_1, a_2, \dots, a_n can be imbedded in E_{n-1} so that $T_0 \simeq T'_0$ in one of two ways: (1) v' on the same side of E_{n-1} as a'_1 ; (2) v' on the opposite side.

In the first case, since v lies within T_1 , the congruence $v + a_2 + \dots + a_n \simeq v' + a'_2 + \dots + a'_n$, which is a sub-congruence of $v + a_1 + \dots + a_n \simeq v' + a'_1 + \dots + a'_n$, defined in the preceding paragraph, is also a sub-congruence of $T_1 \simeq T'_1$, which includes $T_{01} \simeq T'_{01}$.‡ If $x \sim x'$ in the congruence $T_1 \simeq T'_1$, we have $a_0 + v + x \simeq a'_0 + v' + x'$ and (by the previous paragraph) $v + x + a_1 \simeq v' + x' + a'_1$. Now by the four-point property

* Presented by title to the Society, September, 1934.

† "A relation between metric and Euclidean spaces," *American Journal of Mathematics*, vol. 54 (1932), pp. 505-517.

‡ For we can then refer to Theorem I of § 8 instead of Theorem IV.

$a_0 + v + x + a_1 \approx a''_0 + v' + x' + a'_1$, where a''_0 is some point of E_{n-1} . These congruences combined give $v'a'_0 = va_0 = v'a''_0$ and $x'a'_0 = xa_0 = x'a''_0$, while $x'a'_0 = x'v' + v'a'_0$ and $x'a''_0 = x'v' + v'a''_0$. Hence $a'_0 = a''_0$ and so $a_0a_1 = a'_0a'_1$.

Precisely the same argument applies when v' and a'_1 are on opposite sides of E_{n-2} . Thus the assumption that $\text{sign } D(a_0, a_1, \dots, a_n) = \text{sign } (-1)^n$ is false, since it has led to the contradiction that a_0, a_1, \dots, a_n can be imbedded in E_{n-1} . That is, the theorem in question is valid when external convexity is not given.

It follows, therefore, that the theorem of § 12 (*loc. cit.*) can be modified to read: *A convex complete separable space which has the four-point property is congruent with a sub-set of some E_n or of Hilbert space.**

2. Turning to the second topic, we recall Menger's condition † for imbedding $n + 1$ points of a metric space in E_n , namely that, if the distances between the respective pairs of any $k + 1$ ($k \leq n$) of these points are substituted in the formula for the volume of a k -dimensional simplex in terms of the edges, the result is real. This condition can be put into another form which is of some interest.

Let the $n + 1$ points be designated by the integers $0, 1, 2, \dots, n$; then $01, 02$, etc., will denote segments or lengths of segments. Assuming for the moment that the points can be imbedded in E_n , let $0:rs$ denote the angle between the segments $0r$ and $0s$. Then

$$(1) \quad (rs)^2 = (0r)^2 + (0s)^2 - 2(0r)(0s) \cos 0:rs.$$

For four points $0, 1, r, s$, which are the vertices of a tetrahedron let $01:rs$ denote the dihedral angle of edge 01 and faces $01r$ and $01s$. It is well known that

$$(2) \quad \cos 0:rs = \cos 0:1r \cos 0:1s + \sin 0:1r \sin 0:1s \cos 01:rs.$$

In general, if $0, 1, \dots, k + 1, r, s$, are the vertices of a $k + 3$ dimensional simplex, let $01 \dots k + 1:rs$ denote the space-angle having the "edge" $01 \dots k + 1$ and the "faces" $01 \dots k + 1, r$ and $01 \dots k + 1, s$. (This is the angle between two $k + 2$ -dimensional spaces in a $k + 3$ -dimensional

* The referee states that the reasoning employed above can be applied with little change to the corresponding work of L. Blumenthal, "Concerning spherical spaces," *American Journal of Mathematics*, vol. 57 (1935), pp. 51-61. See Theorems 3.3, 4.1, and 4.2. The property of external convexity corresponds to that of being "diamaterized" in spherical spaces.

† "Untersuchungen über allgemeine Metrik, II," *Mathematische Annalen*, vol. 100, pp. 133 and 136.

space.) The spherical cosine law is also valid for these generalizations of dihedral angles, giving*

$$(3) \quad \cos 01 \cdots k:rs = \cos 01 \cdots k:k+1, r \cos 01 \cdots k:k+1, s \\ + \sin 01 \cdots k:k+1, r \sin 01 \cdots k:k+1, s \cos 01 \cdots k+1:rs.$$

Now, if V is the volume of the simplex $(012 \cdots n)$, it is known† that the formula used by Menger can be transformed with the aid of (1) into

$$V^2 = \frac{(01)^2(01)^2 \cdots (0n)^2}{(n!)^2} \cdot \Delta_n,$$

$$\text{where } \Delta_n = \begin{vmatrix} 1 & \cos 0:12 & \cos 0:13 & \cdots & \cos 0:1n \\ \cos 0:12 & 1 & \cos 0:23 & \cdots & \cos 0:2n \\ \cos 0:13 & \cos 0:23 & 1 & \cdots & \cos 0:3n \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cos 0:1n & \cos 0:2n & \cos 0:3n & \cdots & 1 \end{vmatrix}.$$

Multiply the first column successively by $\cos 0:12$, $\cos 0:13$, etc., and subtract from the columns headed by these factors. Clearly $1 - \cos^2 0:1s = \sin^2 0:1s$. We also get terms of the form $\cos 0:rs - \cos 0:1r \cos 0:1s$. In such cases substitute $\sin 0:1r \sin 0:1s \cos 01:rs$ by means of the spherical cosine law (2). We can then remove common factors and get

$$\Delta_n = \sin^2(0:12) \sin^2(0:13) \cdots \sin^2(0:1n) \cdot \Delta_{n-1},$$

$$\text{where } \Delta_{n-1} = \begin{vmatrix} 1 & \cos 01:23 & \cos 01:24 & \cdots & \cos 01:2n \\ \cos 01:23 & 1 & \cos 01:34 & \cdots & \cos 01:3n \\ \cos 01:24 & \cos 01:34 & 1 & \cdots & \cos 01:4n \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cos 01:2n & \cos 01:3n & \cos 01:4n & \cdots & 1 \end{vmatrix}.$$

We treat this determinant as we did Δ_n , using formula (3) above for the case that $k = 1$ and we get

$$\Delta_{n-1} = \sin^2(01:23) \sin^2(01:24) \cdots \sin^2(01:2n) \cdot \Delta_{n-2},$$

* This is given by James McMahon, "Hyperspherical goniometry and its application to correlation theory for n variables," *Biometrika*, vol. 15 (1923), p. 187. It can also be deduced from a result of Ernst Liers, "Über den Inhalt des vier dimensionalen Pentaeders," *Archiv der Mathematik und Physik*, 2d Series, vol. 12, pp. 344-351.

† See Study, *Zeitschrift für Mathematik und Physik*, vol. 27, p. 150.

where Δ_{n-2} has $n-2$ rows and columns and its elements are cosines of the space-angles $012:rs$.

Continuing in the same fashion, we finally reach the relation

$$\Delta_4 = \sin^2(012 \cdots n-4:n-3, n-2) \sin^2(012 \cdots n-4:n-3, n-1) \\ \sin^2(012 \cdots n-4:n-3, n) \cdot \Delta_3,$$

where

$$\Delta_3 = \begin{vmatrix} 1 & \cos 012 \cdots n-3:n-2, n-1 & \cos 012 \cdots n-3:n-2, n \\ \cos 012 \cdots n-3:n-2, n-1 & 1 & \cos 012 \cdots n-3:n-1, n \\ \cos 012 \cdots n-3:n-2, n & \cos 012 \cdots n-3:n-1, n & 1 \end{vmatrix} \\ = \sin^2(012 \cdots n-3:n-2, n-1) \sin^2(012 \cdots n-3:n-2, n) \sin^2(012 \cdots n-2:n-1, n).$$

This reduction expresses V^2 as the product of non-negative factors. It follows, then, for $n+1$ points in any metric space, that V is real if formulas (1), (2), and (3) above define real angles and the reduction can be carried to the end. Looking back, we observe that the angles $0:rs$ defined by the plane cosine law always have definite real values, since the space is metric. The successive space-angles $012 \cdots k+1:rs$ ($k=0, 1, \cdots, n-3$), are defined by the spherical cosine law (3), which can be written

$$\cos 012 \cdots k+1:rs \\ = \frac{\cos 012 \cdots k:rs - \cos 012 \cdots k:k+1, r \cos 012 \cdots k:k+1, s}{\sin 012 \cdots k:k+1, r \sin 012 \cdots k:k+1, s}.$$

The space-angle thus defined has a definite real value unless the absolute value of the fraction is greater than 1 or the denominator is zero.

Let us now assume the following postulates for angles of all orders:

- I. $012 \cdots k:rs + 012 \cdots k:rt + 012 \cdots k:st \leq 2\pi$;
- II. $|012 \cdots k:rs - 012 \cdots k:rt| \leq 012 \cdots k:st \leq 012 \cdots k:rs + 012 \cdots k:rt$.

If $\sin 012 \cdots k:k+1, r = 0$, this angle is 0 or π . It then follows from these postulates that angles $012 \cdots k:k+1, t$ and $012 \cdots k:rt$ are equal or supplementary for every value of t . In that event the columns of Δ_{n-k} which contain $\cos 012 \cdots k:k+1, r$ are identical or one is the negative of the other. In both cases $\Delta_{n-k} = 0$ and the introduction of higher space-angles is unnecessary, as $V^2 = 0$.

If neither $\sin 012 \cdots k:k+1, r$ nor $\sin 012 \cdots k:k+1, s$ is zero, we can easily show by elementary trigonometry that $1 + \cos 012 \cdots k+1:rs \geq 0$ and $1 - \cos 012 \cdots k+1:rs \geq 0$; whence $012 \cdots k+1:rs$ has a definite real value between 0 and π . Thus, if the above angle postulates hold for

every k , either $V^2 = 0$ or the reduction of the determinants can be carried out to the end.

We can then state as a theorem that $n + 1$ given points of a metric space can be imbedded in E_n unless there is some set of $k + 3$ points, $1 \leq k \leq n - 2$, determining three angles or space-angles $a_1 a_2 \cdots a_k : a_r a_s$, $a_1 a_2 \cdots a_k : a_r a_t$, $a_1 a_2 \cdots a_k : a_s a_t$ such that their sum is greater than 2π or the metric triangle inequality fails.

The reader will note that this result is an extension of Blumenthal's theorem on the equivalence of the four-point property and Postulates I and II for plane angles.* Neither result is any simpler to apply to a metric space defined by the distances between the respective pairs of points than are Menger's criteria, but both have some interest as showing that the determinant criteria may be regarded as phases of the triangle inequality.

YALE UNIVERSITY,
NEW HAVEN, CONN.

* L. M. Blumenthal, "A note on the four point property," *Bulletin of the American Mathematical Society*, vol. 39, pp. 423-426. It may be remarked in the case of four points that the condition that $0 : ab + 0 : ac + 0 : bc \leq 2\pi$ is unnecessary, as a failure of this at any vertex involves a failure of the metric triangle inequality at some other vertex.

ON SEMICOMPACT SPACES.†

By LEO ZIPPIN.

1. *Introduction.* It is a well established idea in topology to consider spaces in which certain general properties are assumed to hold only locally, and it is not new to go a step further and transfer these properties from neighborhoods to boundaries of neighborhoods. None the less the fundamental notion of compactness does not appear to have been treated from this point of view. It is the object of this paper to prove two theorems, based on a property we call semicompactness, which seem to us not uninteresting.

Definition. A topologic (Hausdorff) space C is called *semicompact at a point* x if every neighborhood U_x contains a V_x such that $B(V_x)$, the boundary of V_x , is compact. It is called *semicompact* if it has this property at every point. Of course $B(V_x)$ is necessarily closed so that it is actually self-compact.

1.1. Now we have allowed that $B(V_x)$ may be vacuous. Therefore it is clear, for example, that every zero-dimensional topologic space is semicompact. This suggests, at least, that this concept is hardly likely to be very fruitful without some restriction on the nature of the topologic space C . In this paper we shall go quite a way in delimiting the class of spaces we consider. We shall require of C that it be a separable, complete metric space; that is to say that C can be metrized in such a way that every Cauchy sequence converges. In these spaces it will transpire that the notion of semicompactness is strikingly near to that of local compactness. A semicompact C which is separable and complete metric will be called, for shortness, an s. C.-space.

1.2. The two theorems of this paper are concerned with the possible compactifications of an s. C.-space. Thus while a locally compact separable metric space can be compactified by the addition of a single "point," an s. C.-space may always be compactified by the addition of a countable set (Theorem I). In Theorem II we demand that the s. C.-space be connected and locally connected and obtain a considerable generalization of a Theorem of Freudenthal.‡ We shall conclude with a few remarks on, and some applications of, this theorem.

† Presented to the American Mathematical Society December, 1933. See Abstract, *Bulletin of the American Mathematical Society*, vol. 40 (1934), p. 56, no. 97.

‡ See § 6 and note thereto. We were not aware of Freudenthal's paper at the time of publication of the Abstract for this paper.

2. THEOREM I. Every $s.C$ -space C may be compactified by the addition of a countable point-set.

We may suppose that C is not compact, otherwise the theorem is trivial. We shall associate with C a metric in which every Cauchy sequence converges.

Definition. An open subset V of C such that $B(V)$, the boundary of V , is compact will be called a *domain*: an ϵ -domain if its diameter $< \epsilon$.†

2.1. *The ϵ -Partition.* Let V denote any non-compact domain of C , e. g. C itself. Then \bar{V} is complete in our chosen metric and there must exist an $\epsilon' > 0$ such that \bar{V} is not the sum of any finite number of subsets of diameter $< \epsilon'$.‡ Any positive number $< \epsilon'$ will be called suitable for our Partition. Choose some fixed ϵ , $0 < \epsilon < \epsilon'$. From the separability of C and its semi-compactness, there exists a sequence of ϵ -domains, U_1, U_2, \dots , which cover \bar{V} . Let $K_1 = V \cdot U_1$ and, generally, $K_n = V \cdot (U_n - \sum_{i=1}^{n-1} \bar{U}_i)$.

2.2. We assert that the K_n are ϵ -domains.§ It is obvious that they are small enough, and open. We must show that $B(K_n)$ is compact. It is if it is vacuous. If it is not vacuous let $x \subset B(K_n) \subset \bar{K}_n \subset \bar{V} \cdot \bar{U}_n$.¶ Then $x \not\subset \sum_{i=1}^{n-1} \bar{U}_i$ which is open and contains no point of K_n . Then if $x \subset \sum_{i=1}^{n-1} \bar{U}_i$, $x \subset \bar{U}_k - U_k = B(U_k)$ for some $k < n$. On the other hand if $x \not\subset \sum_{i=1}^{n-1} \bar{U}_i$, then either $x \not\subset U_n$ or $x \not\subset V$. For if $x \subset V \cdot U_n$, it follows that $x \subset K_n$: by assumption, however, $x \subset B(K_n)$. Therefore since $x \subset \bar{V} \cdot \bar{U}_n$, it follows that $x \subset \bar{V} - V = B(V)$, or else $x \subset \bar{U}_n - U_n = B(U_n)$. Then $B(K_n) \subset B(V) + \sum_{i=1}^n B(U_i)$, and this sum is compact.

2.3. It is important for us to notice that although the \bar{K}_n are not open they "cover" \bar{V} in a very definite sense. Let z be any point of \bar{V} and n the least integer such that $z \subset U_n$. Let z_1, z_2, \dots , be any sequence of points of \bar{V} converging to z . Without loss of generality we may suppose them in U_n . Let y denote an arbitrary one of the points z, z_1, z_2, \dots , and let y_1, y_2, \dots , denote a sequence of points of V converging to y . We may assume that these points

† This is a slight departure from customary terminology, which we emphasize by italics.

‡ Otherwise \bar{V} would be compact. See Hausdorff, *Mengenlehre*, 2nd Ed., p. 108.

§ We agree that the null-set is open, therefore an ϵ -domain.

¶ Here, as in the sequel, x denotes any point of the space, restricted in so far only as is immediately made evident.

also are in U_n . Now let x denote an arbitrary one of the points y_1, y_2, \dots , and let m be the *least* integer such that $x \in \bar{U}_m$. This integer depends on x , of course, but for every choice of x , $m \leq n$. Finally, let x_1, x_2, \dots , denote a sequence of points of U_m converging to x . All but a finite number of these are in V , and at most a finite number of them can belong to $\sum_{i=1}^{m-1} \bar{U}_i$. Therefore almost all of them belong to K_m and consequently $x \in \bar{K}_m$. Then it is clear that for every integer k , $y_k \in \sum_{i=1}^m \bar{K}_i$. But then y too belongs to this set, and this means that every one of the points z, z_1, z_2, \dots , belongs to it. What we have proved can be expressed as follows: to every point z of \bar{V} there exists an n such that z is an inner point (relative to \bar{V}) of $\sum_{i=1}^n \bar{K}_i$.

2.4. Let $O_n = \bar{V} - \sum_{i=1}^n \bar{K}_i$. It is clear that the O_n form a monotonic sequence, i. e. $O_n \supset O_{n+1}$, whose product is vacuous, and that each is open relative to \bar{V} . It is easily seen † that $B(O_n)$ is compact. Now since $B(V)$ is compact and closed, there is an integer n such that $B(V)$ is a subset of inner points, relative to \bar{V} , of $\sum_{i=1}^n \bar{K}_i$, i. e. no point of $B(V)$ is a limit point of $\bar{V} - \sum_{i=1}^n \bar{K}_i = O_n$. This follows from the concluding remark of the previous section by an application of the Heine-Borel Theorem. Now $B(V) \cdot \bar{O}_n = 0$ implies $\bar{O}_n \subset V$. Let D_1 denote the first O_m such that $\bar{D}_1 \subset V$. Then D_1 is a *domain* of C . Let D_2 denote the first O_m thereafter, such that $\bar{D}_2 \subset D_1$. It is clear that we can find a subsequence D_m of the O_n such that:

$$\bar{V} \supset V \supset \bar{D}_1 \supset D_1 \supset \bar{D}_2 \supset D_2 \supset \dots$$

It is clear that each D_n is a *domain* of C and that $\Pi \bar{D}_n = 0$. Notice that no D_n is vacuous, since $\bar{V} \not\subset \sum_{i=1}^N \bar{K}_i$ for any integer N , by our choice of ϵ . The sequence of *cells*, K_n , will be called an ϵ -partition of \bar{V} . The corresponding sequence D_m will be said to define an *ideal point* associated with this partition.

3. *The ideal points of C.* For notation's sake, we write $C = C_1^0$. Let us make an ϵ_0 -partition of C_1^0 for a suitable ‡ $\epsilon_0 < 1$. We designate by P_1^0 the associated ideal point and by $D_{1,m}^0$, ($m = 1, 2, \dots$), the *domains* defining P_1^0 . Now let C_1^1, C_2^1, \dots , denote those of the cells of this partition which

† Compare § 2.2.

‡ See § 2.1.

are not compact. If there are any, we make an ϵ_{1n} -partition of each C_n^1 for a suitable ϵ_{1n} (which varies with the cell) $< \frac{1}{2}$. This is possible since each cell is a *domain*. The associated ideal point is denoted by P_n^1 , its defining *domains* by $D_{n,m}^1$, ($m = 1, 2, \dots$). Now let C_1^2, C_2^2, \dots , denote those cells which are not compact which result from any one of the countable set of preceding partitions: their totality is at most countable. Each of these, if any exist, is subjected to an $\epsilon_{2,n}$ -partition, every $\epsilon_{2,n} < 1/4$.

Now we may arrive at an integer N such that all the cells confronting us after the N -th partition are compact.† In this event the process will be terminated and no further ideal points introduced. Otherwise we continue the partitioning indefinitely, every non-compact cell C_m^N of the N -th stage being $\epsilon_{N,m}$ -partitioned for a suitable $\epsilon_{N,m}$ (depending on the cell) $< 1/2^N$, ($N = 1, 2, 3, \dots$).

Whichever of the above alternatives we face, it is clear that we have introduced an at most countable set P of ideal points where each one is some P_m^N in our construction,‡ being associated with a cell C_m^N , $\text{diam.}(C_m^N) < 1/2^N$. The point P_m^N is defined by a properly monotonic sequence, $D_{m,n}^N$, ($n = 1, 2, \dots$), of domains of C_m^N , $\prod_{n=1}^{\infty} \bar{D}_{m,n}^N = 0$.

3.1. Now let C'' denote the abstract "point-set" $C + P$ topologized as follows. Let G_1, G_2, \dots , denote a sequence of *domains* of C which generates § the space C and which includes every defining *domain* for every ideal point P_m^N in P . Now if P_m^N is any ideal point of C , and G_k any domain of the sequence, we shall say that P_m^N belongs to G_k if this set contains any one of the defining *domains* for P_m^N (in which case, of course, it contains almost all of them). Now let G_1'', G_2'', \dots , denote the point sets G_n to which have been added all the ideal points belonging to them. By definition, each G_n'' is a neighborhood of every point of C'' which it contains. Let us observe at once that $G_m \cdot G_n = 0$ (in C) implies $G_m'' \cdot G_n'' = 0$ (in C''). This is obvious, for if $G_m'' \cdot G_n''$ contained an ideal point it would have to contain a non-vacuous *domain* of C , and if it contained a point of C this would have to be a point of $G_m \cdot G_n$.

3.2. It is trivial that every point of C'' belongs to at least one neighborhood of the system and that if it belongs to two neighborhoods it must belong

† Actually this cannot happen unless $C_1^0 = C$ is locally compact, but that is immaterial to the proof.

‡ The ranges of N and of m in its dependence on N depend on the particular choice of partitions.

§ i. e. is a *basis* for the neighborhoods of C .

to a neighborhood common to both of them. Then, since each G_n'' is a neighborhood of every one of its points, we merely have to show that if x and y are distinct points of C'' , there exist $G_m'' \supset x$, $G_n'' \supset y$, $G_m'' \cdot G_n'' = 0$, in order to conclude that C'' is a Hausdorff space. This is trivial, excepting possibly in the case that x is an ideal point P_k^N and y is some $P_{k'}^{N'}$ of P . Here we may suppose, on symmetry, that $N \leq N'$. Now P_k^N is associated with the partition of a non-compact cell C_k^N and $P_{k'}^{N'}$ with that of $C_{k'}^{N'}$. If $N' = N$, $k' = k$, the two points are not distinct. If $N' = N$, $k' \neq k$, P_k^N belongs to a domain $G_m \subset C_k^N$ and $P_{k'}^N$ belongs to a domain $G_n \subset C_{k'}^N$ and $G_m \cdot G_n \subset C_k^N \cdot C_{k'}^N = 0$. On the other hand, if $N' > N$, $C_{k'}^{N'} \subset C_h^N$ for some h . If $h \neq k$ we have the same situation as above. If $h = k$ then $C_{k'}^{N'} = C_j^{N+1} \subset C_k^N$. But then there is a domain G_m to which P_k^N belongs such that $G_m \cdot C_j^{N+1} = 0$ and $P_{k'}^{N'}$ belongs to a subdomain G_n of C_j^{N+1} . Therefore, in view of the last remark of § 3.1, we have been able, whichever of the cases above may have arisen, to find $G_m'' \supset P_k^N = x$ and $G_n'' \supset P_{k'}^{N'} = y$ such that $G_m'' \cdot G_n'' = 0$. Therefore C'' is certainly a Hausdorff space. It is trivial that C'' is completely separable (i. e. has a countable neighborhood basis). It is clear that C may now be regarded as a topologic subspace of C'' , if we ignore the convenient metric we have attached to it.

3.3. Let us prove finally, that C'' is compact.† To this end, let x_1, x_2, \dots denote any sequence of points of C'' .

i) If there exists any integer N such that infinitely many of the cells resulting from the first N partitions contain at least one point of the sequence, then there is a first such N . Then the sequence (x_n) has an infinite subsequence in some C_m^{N-1} and has at least one point in common with every $D_{m,n}^{N-1}$, ($n = 1, 2, \dots$). Therefore, in this case, the ideal point P_m^{N-1} is a limit point (not necessarily the only one) of the sequence (x_n) .

ii) If there is any N such that an infinite subsequence of (x_n) belong to a compact cell (in C) or belong to the boundary (which is compact) of any cell of the N -th partitions, then the subsequence consists of points of C which have at least one limit point in C and this is also a limit point of the given subsequence, in C'' .

iii) Finally, if neither of the previous cases ever arises, it is easy to see that we can find a monotonic sequence of non-compact cells,

$$C = C_1^0, C_{n_1}^1, C_{n_2}^2, \dots,$$

such that for every $C_{n_m}^m$ there is at least one point x_{k_m} of our sequence which

† We may suppose all topologic notions defined for C'' , as customarily.

belongs to it. But by our construction, the diameters of these cells converge to zero. Then it is a well known consequence of the completeness of C (our metric exhibiting this completeness) that there is a unique point of C common to the closures of these cells, and it is clear that this point is a limit point in C'' of the given sequence.

Therefore C'' is a compact, completely separable Hausdorff space and, as is well known, metrizable. We have observed that our space C is topologically equivalent to a subset C' of $C'' = C + P$ where P is countable and $P \subset \bar{C} = C''$. Then C'' is a compactification of C and Theorem I is proved.

4. We may remark that Theorem I is characteristic of s. C.-spaces. This follows from the simple

THEOREM. *If C'' is any compact, metrizable space and $C = C'' - Q$ where Q is any totally disconnected F_σ ,† then C is an s. C.-space.*

It is clear that C must be separable metric, and well known (Alexandroff) that it is complete in some metric. We merely have to show that it is semi-compact. This will follow if we can show that, under our hypotheses, every point of C'' has arbitrarily small neighborhoods whose boundaries are vacuous relative to Q . Write $Q = \sum Q_n$, where Q_n is closed, and totally disconnected. Therefore Q_n is zero dimensional in the Menger-Urysohn sense.

4.1. Then if x denotes any point of C'' , $x + Q$ is a zero dimensional point-set.‡ Therefore, for any fixed $\epsilon > 0$ we can write $x + Q = H_1 + H_2$, $H_1 \cdot H_2 = 0$, where $H_1 \supset x$, $\text{diam.}(H_1) < \epsilon/3$, and both sets are closed in $x + Q$. Now cover every point y of H_1 by an open set D_y (of C'') $0 < \text{diam.}(D_y) < \text{Min.}[1/3\epsilon, 1/2 \text{ dist.}(y, H_2)]$, and let $D = \sum_y D_y$. It is clear that $0 < \text{diam.}(D) < \epsilon$, and that $H_1 \subset D$ which is open. It is easy to see that $H_2 \cdot \bar{D} = 0$.§ Then the boundary of D cannot contain any point of Q so that D is the desired neighborhood, and the theorem is proved.

We need hardly remark that it is not necessary that an F_σ subset Q of a compact metrizable C'' be totally disconnected in order that $C = C'' - Q$ shall be an s. C.-space.

† $Q = \sum Q_n$, Q_n closed. This includes the case that Q is countable.

‡ We are appealing to the "Summensatz" of dimension-theory. A proof of what we need can be carried through by a method which Menger has called "Methode der Modification der Umgebungen in der Nähe ihrer Begrenzungen" and on which his proof of the Summensatz rests. See his book *Dimensionstheorie*, p. 94.

§ Compare the lemma of Urysohn, "Sur les multiplicités Cantoriniennes," *Fundamenta Mathematicae*, vol. 7 (1925), p. 69.

5. It is clear that if an s. C.-space C is connected, the C'' of Theorem I is also connected. If, further, C is locally connected, we may suppose that those G_n (of § 3.1) which generate C were chosen as connected point-sets and the corresponding G_n'' will be connected. Then, in this case, C'' will certainly be locally connected at every point of C . Consequently, by a theorem of Mazurkiewicz, C'' will be locally connected since $C'' - C = P$ is totally disconnected.

Definition. A connected and locally connected s. C.-space will be called *semipeanian*.†

5.1. We have just proved the

COROLLARY. A *semipeanian* C is topologically contained in a *peanian* $C'' = \bar{C} = C + P$, P countable.

Now there are many possible compactifications of C . If we require that the set of ideal points which we adjoin shall be at most countable, then there is not any C'' which is invariantly associated with a general C .‡ Moreover, in this case, the ideal points of C' will, in general, "interrupt" C'' , in the sense, for example, that it may not be possible to join two neighboring points of C by a small arc of C'' which avoids P . If we do not insist on compactifying C with a countable point-set, then we can show that there exists a *peanian* C^* invariantly associated with C and rather simply related to it. We may say that the ideal points offer a minimum of interruption. The sense of this will be made precise in Theorem II.

Definition. A totally disconnected subset Q of a *peanian* C^* will be called totally avoidable provided that $D - D \cdot Q$ is connected for every open connected subset D of C^* .§

6. THEOREM II.¶ *Every semipeanian C is topologically contained in a*

† Complete-metric, separable, connected and locally connected spaces are commonly called *quasipeanian*. Thus, *semipeanian* = *quasipeanian* + *semicompact*. Compact, metr., con. and loc. con. spaces we shall call *peanian*.

‡ We shall return to this in § 7.

§ This is a special case of a more general definition of total avoidability, due to Wilder.

¶ We have already remarked that this is a Theorem of Freudenthal in the case that C is locally compact. See H. Freudenthal, "Über die Enden topologischer Räume und Gruppen," *Mathematische Zeitschrift*, vol. 33 (1931). Satz 7, p. 702. A similar compactification was used by us in characterizing subsets of a simple closed surface which we called *cylinder-trees*. See "Study of continuous curves . . .," *Transactions of the American Mathematical Society*, vol. 31 (1929), Theorem 6, p. 763. However

uniquely determined peanian $C^* = \bar{C}$ such that $Q = C^* - C$ is a totally disconnected and totally avoidable F_σ .

Proof. By the corollary of § 5.7, C may be compactified to a peanian $C'' = C + P$, P countable.† Let U_n'' , ($n = 1, 2, \dots$), be a null-sequence‡ of open connected subsets generating C'' such that $P \cdot B(U_n'') = \emptyset$.§ If U'' denotes any U_n'' , $U = C \cdot U''$, and x is any point of U , then $U \supset U_x \supset x$ where U_x is connected and open in C . This is an immediate consequence of the fact that C is topologically contained in C'' , and is locally connected. It follows at once that the set of components of U is at most countable.

6.1. Although we do not need it at this moment it is convenient to prove now that if p is any point of $P \cdot U''$ and $2\epsilon = \text{dist.}\{p, B(U'')\}$ there are only a finite number of components of $U = C \cdot U''$ which meet $S(p, \epsilon)$.¶ For if x is any point of $U \cdot S(p, \epsilon)$ any y denotes any point of $C - U$, the existence of an arc xy of C shows that the component $U_x \supset x$, of U , has at least one point on the boundary of every $S(p, \epsilon')$ where $\epsilon < \epsilon' < 2\epsilon$. Then if there were infinitely many components in question, there would exist at least one point $x_{\epsilon'}$ on $B\{S(p, \epsilon')\}$ which was a limit point of points of distinct components. Now $x_{\epsilon'} \not\subset U$, since the components are open in U . Therefore $x_{\epsilon'} \subset P$. But this is impossible since the $x_{\epsilon'}$ are distinct for different ϵ' , and P is at most countable.

6.2. *Ideal points of C .* The totality of components of $C \cdot U_n''$, ($n = 1, 2, \dots$), is countable, by the last remark of § 6. We denote them, in some simple order by W_1, W_2, \dots . A monotonic sequence W_{n_i} , ($i = 1, 2, \dots$), of sets W_m will be called a *proper sequence* if the product of their closures (in C'') is a single point of P . Two proper sequences W_{n_i}, W_{m_i} , ($i = 1, 2, \dots$), are called *equivalent* if for every j there is a k such that $W_{n_j} \supset W_{m_k}$ and conversely to every k a j such that $W_{m_k} \supset W_{n_j}$. It is trivial that our definition satisfies the usual conditions for equivalence. A class of equivalent proper sequences will be called an *ideal point* of C . It is clear that with each ideal point of C there is associated a unique point of P , this correspondence being, in general, many-one. The totality of ideal points we denote by Q . We shall

this process is there carried out in a very special case and its essential generality was not then suspected by us. Our method there, as here, differs from Herr Freudenthal's in that we exploit a preliminary compactification of the space.

† The use of C'' is a pure convenience to facilitate the handling of the ideal points which we presently define.

‡ i. e. $\text{diam.}(U_n'')$ converges to zero in an arbitrary fixed metric for C'' .

§ This condition is easily fulfilled. See § 4.1.

¶ The set of points whose distance from p is $< \epsilon$.

say that an ideal point q belongs to a set W of C if $W \supset W_{n_k}$, where W_{n_k} is any set in any proper sequence defining q . It is clear that W contains almost all the sets in any equivalent proper sequence. Observe that if q is an ideal point, p the associated point of C'' , W'' any neighborhood of p in C'' and $W = C \cdot W''$, then q belongs to W .

6.3. *The space C^* .* Let C^* denote the abstract point set consisting of points and ideal points of C . We may write this $C^* = C + Q$. Let W_n^* denote the subset of C^* consisting of all points of $W_n \subset C$ and all points of Q which belong to W_n by the definition of the preceding section. We shall topologize C^* by agreeing that W_n^* , ($n = 1, 2, \dots$), is a neighborhood of every one of its points. We observe that $W_m \cdot W_n = 0$ (in C) implies $W_m^* \cdot W_n^* = 0$ (in C^*).† It is trivial that these neighborhoods have all the Hausdorff properties with the possible exception of this one: that if x^* and y^* are distinct points of C^* there exist $W_m^* \supset x^*$, $W_n^* \supset y^*$, $W_m^* \cdot W_n^* = 0$. This is also trivial in the case that the points x and y of C'' associated‡ with x^* and y^* are distinct, in view of the observations above. We shall dispose of the remaining case in § 6.5.

6.4. Let us suppose that q is an ideal point of C and that every W_{n_j} , ($j = 1, 2, \dots$), of any corresponding proper sequence intersects a fixed W_n . We shall prove that q belongs to W_n . Let p denote the associated point of $P \subset C''$. Now $p \subset \bar{W}_n$ (in C''). For otherwise there is a neighborhood D'' of p , $D'' \cdot W_n = 0$. We have already observed that q must belong to $D = C \cdot D''$ so that for some j , $W_{n_j} \subset D$ and $W_{n_j} \cdot W_n = 0$ which is contrary to assumption. Therefore $p \subset \bar{W}_n \subset \bar{U}_m''$ for that m for which the given W_n is a component of $C \cdot U_m''$. Then $p \subset U_m''$, since $B(U_m'') \cdot P = 0$ by construction.§ Now if we consider the sets U_i'' which correspond to the W_{n_j} , it is clear that there must occur among them sets U_i'' of indefinitely large subscript, and therefore of arbitrarily small diameter since the U_i'' form a null-sequence. Otherwise it would follow that there were only a finite number of distinct W_{n_j} , and this would imply by the monotonic character of these sets that for some k , $W_{n_k} \subset W_{n_j}$ for every j . But then $p = \bigcap_{j=1}^{\infty} \bar{W}_{n_j} = \bar{W}_{n_k} \supset W_{n_k}$, although $W_{n_k} \subset C$ and is not vacuous. This is absurd. Then, since $p \subset U_n''$ there is a W_{n_j} such that the corresponding $U_i'' \subset U_m''$. It follows, exactly as above, that $p \subset U_i''$. Now $W_{n_j} \subset U_m'' - P \cdot U_m''$ and is connected. Further

† Compare § 3.7, last remark.

‡ If $x^* \subset C$, $x = x^*$.

§ See § 6.2.

$W_{n_j} \cdot W_n \neq 0$ and W_n is a component of $U_m'' - P : U_m''$. Therefore $W_{n_j} \subset W_n$ and therefore q belongs to W_n .

6.5. Now, to return to the argument of § 6.3, let us suppose that x^* and y^* are points of $Q \subset C^*$ such that $W_m^* \supset x^*$, $W_n^* \supset y^*$ implies $W_m^* \cdot W_n^* \neq 0$, therefore $W_m \cdot W_n \neq 0$. Let W_{m_i} and W_{n_i} , ($i = 1, 2, \dots$), define the ideal points x^* and y^* . Then $W_{m_j} \cdot W_{n_k} \neq 0$ for every j and k . If we keep j fixed but $k = 1, 2, \dots$, we see from the previous section that almost all the $W_{n_k} \subset W_{m_j}$. Keeping k fixed, but letting $j = 1, 2, \dots$, we see that almost all the $W_{m_j} \subset W_{n_k}$. Then the two sequences are equivalent and define the same ideal point: i. e. $x^* = y^*$. This concludes the argument that C^* is a Hausdorff space. It is trivial that C^* is completely separable. It is clear, also, that C is topologically contained in C^* , and that every point of Q is a limit point (in the topology of C^*) of points of C . Then $C^* = \bar{C}$ and is connected and every $W_n^* = \bar{W}_n$ which is connected, so that C^* is locally connected; where closure is to be understood in the sense of the topology of C^* . To show that C^* is peanian we merely have to prove that it is compact.

To this end let x_1^*, x_2^*, \dots be any sequence of points of C^* , and x_1'', x_2'', \dots the corresponding sequence of not necessarily distinct associated points of C'' (if $x_n'' \subset C$, $x_n'' = x_n^*$). We may suppose that the second sequence converges to a point x'' of C'' (if $x_m'' = x_n''$ for some m and infinitely many n , then $x'' = x_m''$).

i) $x'' \subset C \subset C^*$. Let W_n^* be any neighborhood of $x^* = x''$ in C^* . Then $x^* \subset W_n \subset C$. Since W_n is open in C there is a neighborhood U'' of x^* in C'' such that $C \cdot U'' \subset W_n$. Almost all the $x_m'' \subset U''$. If $x_m'' \subset P$, for some m , x_m^* belongs to at least one $W_k \subset C \cdot U'' \subset W_n$. Therefore x_m^* belongs to W_n and $x_m^* \subset W_n^*$. If $x_m'' \subset C$, $x_m^* = x_m'' \subset W_n \subset W_n^*$. Then it follows that x^* is a limit point in C^* of the sequence x_1^*, x_2^*, \dots .

ii) $x'' \subset P$. Let U_{n_i}'' , ($i = 1, 2, \dots$), be a monotonic sequence of the neighborhoods generating C'' such that $\cap U_{n_i}'' = x''$. By § 6.7, there is a U_{n_k}'' , $x'' \subset U_{n_k}'' \subset U_{n_1}''$ such that the points of $C \cdot U_{n_k}''$ are contained in the sum of a finite number of the components of $C \cdot U_{n_1}''$. Now almost all the $x_m'' \subset U_{n_1}''$. Therefore there is at least one component W_{n_1} of $C \cdot U_{n_1}''$ such that infinitely many of the points x_m^* belong to W_{n_1} and are contained, therefore, in $W_{n_1}^*$. Then it is possible by an easy "diagonalizing" process to find a subsequence

$$x_{i_1}^*, x_{i_2}^*, \dots, x_{i_n}^*, \dots,$$

of our given sequence of points of C^* , and a monotonic sequence of neighborhoods

$$W_{j_1}^*, W_{j_2}^*, \dots, W_{j_n}^*, \dots,$$

such that each $W^*_{j_n}$ contains almost all of the points of the sequence and such that each U_{n_i}'' of this paragraph contains almost all of the $W_{j_n} = C \cdot W^*_{j_n}$ ($n = 1, 2, \dots$). Then the W_{j_n} ($n = 1, 2, \dots$), form a proper sequence associated with the point x'' of C'' and define an ideal point $x^* \in C^*$. It is clear that $x^* \in W^*_{j_n}$ for every n . Now *every* neighborhood W^*_k of x^* must contain at least one W_{j_n} and therefore the corresponding $W^*_{j_n}$. Then, finally, x^* is a limit point in C^* of our given sequence.

Then we have shown that C^* is a peanian space, and that C is topologically contained in it.

6.6. We shall now consider the point-set $Q \subset C^*$. Let the points of $P \subset C''$ be enumerated in a sequence, p_1, p_2, \dots , and let Q_n be the subset of points of Q associated with p_n , ($n = 1, 2, \dots$). Then the argument we have just given above shows that Q_n is closed (in C^*). Therefore $Q = \sum Q_n$ is an F_σ -set. This is also an obvious consequence of the known *absolute G_δ -character* of the space C . However, the relation of the sets Q and P is not uninteresting.† Let us now show that Q is totally disconnected. This will follow at once when we have shown that the boundaries of our neighborhoods W^*_n are vacuous relative to Q . Now this is merely a restatement of § 6.4. For if a point $q \in B(W^*_n)$, every neighborhood $W^*_{n_j}$ of q contains points of W^*_n . If $q \in Q$ it is an ideal point of C . If W_{n_j} , ($j = 1, 2, \dots$), defines q then $W_{n_j} \cdot W_n \neq 0$ and, by § 6.4, almost every $W_{n_j} \subset W_n$. Then $q \in W^*_n$ and $q \in B(W^*_n)$. We shall show, finally, that Q is totally avoidable in C^* . Let D^* be any open connected subset of C^* , and suppose that x and y are points of $C \cdot D^*$ which belong to no connected subset of $D^* - Q \cdot D^* = C \cdot D^*$. Now D^* is a locally compact peanian space ‡ and it is known that there must exist a point q of Q such that $W^*_n - Q \cdot W^*_n = C \cdot W^*_n = W_n$ is not connected for every neighborhood W^*_n of q . But this is absurd since every W_n is a connected subset of C by construction. Now since an open connected subset of C is necessarily arcwise connected, the argument shows also that if xy is any arc of C^* , $x + y \subset C$, then there is another arc xy of C in every neighborhood (in C^*) of the given arc.

6.7. To finish the proof of our Theorem we must show that C^* is uniquely defined by its relation to C . This includes the statement that C^* is a topological invariant of C . We shall prove somewhat more, namely that if C_1 and C_2 are homeomorphic semi-peanian spaces, $C^*_1 = C_1 + Q_1$, and $C^*_2 = C_2 + Q_2$ the corresponding compactifications with the properties we

† See § 7.

‡ It makes a pretty terminological sequence to call such spaces *near-peanian*.

have already established, and $T(C_1) = C_2$ any homeomorphism carrying C_1 into C_2 then T can be extended to a homeomorphism T^* , $T^*(C_1^*) = C_2^*$, $T^*(C_1) = T(C_1)$.

By a theorem of Alexandroff it will be sufficient to show that T and its inverse are uniformly continuous, since C_1 and C_2 are dense in C_1^* and C_2^* . By argument of symmetry, it is sufficient to prove this for T . Now to do this it is merely necessary to prove that if x_1, x_2, \dots , and x'_1, x'_2, \dots , are two sequences of points of C_1 converging to the same point x of Q_1 , and $y_n = T(x_n)$, $y'_n = T(x'_n)$, then the sequences y_1, y_2, \dots , and y'_1, y'_2, \dots , converge to the same point y of Q_2 . Each of the last two sequences certainly has at least one limit point in C_2^* .

Now if either of these has at least two limit points, or if they do not have the same limit point then we can find a subsequence y_{n_i} , ($i = 1, 2, \dots$), converging to a point y and a subsequence y'_{m_i} , ($i = 1, 2, \dots$), converging to $y' \neq y$. Let x_{n_i} and x'_{m_i} denote the corresponding sequences in C_1 . Since C_1^* is peanian, it contains arcs $x_{n_i}x'_{m_i}$, ($i = 1, 2, \dots$), such that these converge to x , i. e. if $z_i \subset x_{n_i}x'_{m_i}$, then z_i converges to x . Now since Q_1 is totally avoidable, we may suppose without any loss that these arcs belong to C_1 .† Let $y_{n_i}y'_{m_i} = T(x_{n_i}x'_{m_i})$. These arcs belong to C_2 . There is a subsequence of them which converges to a limiting continuum $K \supset y + y'$ of C_2^* . Since Q_2 is totally disconnected, there is at least one point y^* of K , $y^* \subset C_2$, and there is a sequence of points y^*_1, y^*_2, \dots , converging to y^* such that no two of them belong to the same arc $y_{n_i}y'_{m_i}$. Therefore no two of the corresponding points (under the inverse of T) x^*_1, x^*_2, \dots , belong to the same arc $x_{n_i}x'_{m_i}$ and therefore they converge to x . Then the inverse of T cannot be continuous. This contradiction establishes our argument and brings our proof of Theorem II to a close.

7. Here we shall consider the relation of the subset Q of the uniquely defined C^* associated with a semipeanian C and the countable subset P of a compactification C'' . We have seen that if we start with a C'' we arrive at C^* with a resolution of Q into ΣQ_n , where each Q_n is closed, every point of a Q_n is associated ‡ with the same point p_n of $P \subset C''$, and the p_n are distinct for distinct Q_n . Now, conversely, if we consider C^* and write $Q = \Sigma Q_n$, where $Q_m \cdot Q_n = 0$, $m \neq n$, and each Q_n is closed, then each such resolution of Q gives rise to a space C'' . This space C'' is simply the *decomposition space* of C^* where each point of C and each set Q_n is regarded as a point. For it is

† See the last remark of § 6. 6.

‡ See the opening sentences of § 6. 6.

clear that C'' is peanian, since it is the continuous image (when it is topologized as customarily) of the peanian C^* , and contains C topologically as an everywhere dense subset.

8. There is a simple converse to Theorem II.

THEOREM. *If C^* is peanian and Q a totally disconnected and totally avoidable F_σ , then $C = C^* - Q$ is semipeanian.*

We have shown that C is an s. C.-space.† It is clear from the definition of total avoidability that C is connected and locally connected. It need hardly be remarked that it is not necessary that Q be totally disconnected in order that C be semipeanian.

9. *The space I_2 .* The dimension of C^* cannot exceed that of C by more than one, i. e. $\dim C \leq \dim C^* \leq 1 + \dim C$. This is an immediate consequence of the totally disconnected character of $Q = C^* - C$. On the other hand, the dimension of C^* may have the larger value. Thus if C is the space I_2 of irrational points of a Cartesian plane (at least one coördinate irrational) then C^* is a topologic sphere. In this case: $\dim C^* = 2$, $\dim I_2 = 1$. It is amusing that Theorem II permits a characterization of I_2 . It is easy to see that I_2 is 1) semipeanian, 2) nowhere locally compact. It is clear, further, that 3) every simple closed curve J of I_2 separates it and 4) no arc of any J separates I_2 . Finally, if we follow Freudenthal ‡ and define "ends" abstractly as any monotonic sequence of open connected sets D_n , ($n = 1, 2, \dots$), with compact boundaries, such that $\Pi \bar{D}_n = 0$, then 5) I_2 has an at most countable set of *distinct* "ends," distinct being used in the sense of non-equivalent. Although we shall not prove it here it is not difficult to show that these five properties completely characterize I_2 .

10. *Primitive skew curves.* By primitive skew curve we understand either of the two non-planar linear graphs.§

THEOREM. *If C contains no primitive skew curve, then C^* contains none.*

The proof is quite simple. For if K^* is a skew curve of C^* then we can replace each arc of K^* with endpoints in C by an arc of C which lies in an

† Compare § 4.

‡ *Loc. cit.*, p. 695. The *distinct* "ends" coincide with our ideal points Q .

§ See C. Kuratowski, "Sur le probleme des courbes gauches en Topologie," *Fundamenta Mathematicae*, vol. 15 (1930), pp. 271-283.

arbitrary neighborhood of the first.† We can conclude easily that C^* contains a skew curve K'' of exactly the same type as K^* whose vertices, at worst, do not belong to C . It is fairly obvious that if these vertices are of order three we can displace K'' slightly at its vertices and obtain a similar skew curve K in C . If the vertices are of order four we may not be able to "reproduce" K'' in C . None the less it is readily seen that by introducing small arcs of C in the neighborhood of the vertices of K'' we can arrive at a skew curve K of C , which is in general of the first type.‡

The theorem above permits a complete extension to semipeanian spaces of the recent work of S. Claytor.§ This work is a very considerable generalization of a Theorem of Kuratowski ¶ on planar subsets.

11. In large part, it has been the burden of this paper that for quasi-peanian spaces at least, local compactness and semicompactness are very close kin. In this concluding section we shall prove the

THEOREM. *A semipeanian group manifold G has at most two distinct "ends" ¶ in the sense of Freudenthal.*

Let t^* denote any point of $G^* - G$, where G^* is the compactification of G in Theorem II, t_n , ($n = 1, 2, \dots$), a sequence of points of G converging to t^* , and g any element of G . Now each element of G gives rise to a translation of G into itself, which is a homeomorphism. This extends to a unique homeomorphism of G^* into itself where the complement of G is invariant, by § 6. 7. We may denote this extended homeomorphism by g . The translated points $t_n g$ must converge to t^* . For, if they did not, we could find a neighborhood D^* of t^* with boundary in G such that $t_n g \not\subset D^*$ held for infinitely many n ; by thinning our sequence we may say for all n . Now if γ is any arc of G , from g to the identity of G , the translated arcs $t_n \gamma$ must all have at least one point b_n on $B(D^*) \subset G$, where $b_n = t_n a_n$, $a_n \subset \gamma$. We may suppose the

† By the total avoidability of $Q = C^* - C$. See the last remark of § 6. 6.

‡ Compare Mazurkiewicz, "Über nicht plattbare Kurven," *Fundamenta Mathematicae*, vol. 20 (1933), p. 284.

§ I am advised by Claytor that his paper is to appear in the *Annals of Mathematics* in October of this year. See Abstract No. 158, *Bulletin of the American Mathematical Society*, vol. 39 (1933), p. 357.

¶ See note of this section.

* ¶ See note to § 9, also "Satz 15," *loc. cit.* Our argument here will differ very slightly in form but hardly at all in essence from that of Freudenthal. We are obliged to make this change since local compactness is required by one of his subsidiary theorems (Satz 13).

$a_n \rightarrow a \nmid \subset \gamma$, $b_n \rightarrow b \subset B(D^*)$. It follows that the $t_n \rightarrow t$, $t = ba^{-1} \subset G$. Since this is contrary to assumption, the t_n being a "divergent sequence" \dagger in G the assertion is proved. Now this shows that *each point* of $G^* - G$ is invariant under the homeomorphism g , where g is any element of G . \S If we now consider the element g as fixed the t_n as a sequence of homeomorphisms of G^* , then the $t_n g$ now denote the successive translations of g , and these converge to t^* , for every sequence t_n converging to t^* . By uniformity arguments, the translated sets $t_n M$ where M is any self compact subset of G , converge to t^* , so that for an arbitrary open $D^* \supset t^*$, there is an n such that $t_n M \subset D^*$. \P

Let us suppose that $G^* - G$ contains as many as three distinct points x^* , y^* , z^* . Let $V^* \supset x^*$, $W^* \supset y^*$ be neighborhoods, $z^* \not\subset V^* + W^*$, $V^* \cdot W^* = 0$, and $M = B(W^*) \subset G$. It is an easy consequence of the avoidability of x^* that there is a neighborhood $U^* \supset x^*$, $V^* \supset U^*$, such that *any* two points of $G^* - V^*$ can be joined by an arc of $G^* - U^*$: in particular, the point z^* and any other. By the preceding paragraph there is at least one element x of G such that $xM \subset U^*$. Now with every subset H^* of G^* there is associated the homeomorphic set $x\{H^*\}$. Since z^* is a fixed point, $x\{G^* - W^*\} \supset z^*$. Therefore, since

$$B(x\{G^* - W^*\}) = x\{B(G^* - W^*)\} \subset x\{B(W^*)\} = xM \subset U^*,$$

it follows that $x\{G^* - W^*\} \supset G^* - V^*$. From this it must follow that $x\{W^*\} \subset V^*$. Now this is impossible since $y^* \subset W^*$, $y^* \not\subset V^*$ and is a fixed point under the homeomorphism x . Therefore $G^* - G$ cannot consist of more than two distinct points. This shows at once that G must be locally compact and completes the proof.

INSTITUTE FOR ADVANCED STUDY,
PRINCETON, NEW JERSEY.

\dagger Read "converge, as elements of the group manifold G , to": of course, they also converge as points.

\S Associated with the "end" determined by t^* .

\S *Loc. cit.*, "Satz 12."

\P *Loc. cit.*, "Satz 11."

ADDITION THEOREMS FOR THE DOUBLY PERIODIC FUNCTIONS OF THE SECOND KIND.

By WALTER H. GAGE.

1. *Introduction.* In this paper we derive addition theorems for $\phi_{\alpha\beta\gamma}(x, y)$, where

$$\phi_{\alpha\beta\gamma}(x, y) = \frac{\vartheta'_1 \vartheta_\alpha(x+y)}{\vartheta_\beta(x) \vartheta_\gamma(y)},$$

and where ϑ_α ($\alpha = 0, 1, 2, 3$) are the theta functions of Jacobi.* The formulae obtained are addition theorems, not in the ordinary sense, but according to the definition of Poincaré.†

2. *The fundamental formulae.* From the special case of one of Jacobi's theta-identities

$$\begin{aligned} \vartheta_2 \vartheta_1(y+v) \vartheta_3(v+x) \vartheta_0(x+y) \\ = \vartheta_3(x+y+v) \vartheta_0(x) \vartheta_2(y) \vartheta_1(v) + \vartheta_0(x+y+v) \vartheta_3(x) \vartheta_1(y) \vartheta_2(v) \end{aligned}$$

it follows that

$$\begin{aligned} (1) \quad \vartheta_2(y) \vartheta_1(v) \phi_{001}(x, y+v) + \vartheta_1(y) \vartheta_2(v) \phi_{331}(x, y+v) \\ = \frac{\vartheta'_1 \vartheta_2 \vartheta_0(x+v) \vartheta_3(x+y)}{\vartheta_0(x) \vartheta_3(x)}. \end{aligned}$$

If, in (1), we interchange y and v we also have

$$\begin{aligned} (2) \quad \vartheta_1(y) \vartheta_2(v) \phi_{001}(x, y+v) + \vartheta_2(y) \vartheta_1(v) \phi_{331}(x, y+v) \\ = \frac{\vartheta'_1 \vartheta_2 \vartheta_3(x+v) \vartheta_0(x+y)}{\vartheta_0(x) \vartheta_3(x)}. \end{aligned}$$

Solving (1) and (2) for $\phi_{331}(x, y+v)$, and simplifying the result by means of the identity

$$\vartheta_1^2(y) \vartheta_2^2(v) - \vartheta_2^2(y) \vartheta_1^2(v) = \vartheta_2^2 \vartheta_1(y+v) \vartheta_1(y-v),$$

we get

$$\begin{aligned} (3) \quad \phi_{331}(x, y+v) \\ = \frac{\vartheta'_1}{\vartheta_2 \phi_{111}(y, v) \phi_{122}(y, -v)} \{ \phi_{001}(x, v) \phi_{332}(x, y) - \phi_{332}(x, v) \phi_{001}(x, y) \}. \end{aligned}$$

* For the doubly periodic functions see E. T. Bell "Algebraic Arithmetic" page 88; for the definitions and notation of theta functions see Whittaker and Watson "Modern Analysis" Chap. 21 (4th ed.).

† Poincaré, "Sur une Classe Nouvelle de Transcendentes Uniformes," *Journal de Mathématiques*, Quatrième Série, 1890.

Let us write this briefly as

$$(4) \quad (331) = K(111, 122) \{ (001, 332) - (332, 001) \},$$

where

$$K(111, 122) = \vartheta'_1 / \vartheta_2 \cdot \phi_{111}(y, v) \phi_{122}(y, -v).$$

Increasing x by $\pi/2$ gives

$$(5) \quad (001) = K(111, 122) \{ (331, 002) - (002, 331) \}.$$

If we increase x by $\pi\tau/2$ in each of (4) and (5), there results

$$(6) \quad (221) = K(111, 122) \{ (111, 222) - (222, 111) \},$$

$$(7) \quad (111) = K(111, 122) \{ (221, 112) - (112, 221) \},$$

respectively.

The remaining formulae for the sixty triple subscripts $\alpha\beta\gamma$ of $\phi_{\alpha\beta\gamma}(x, y + v)$ can be obtained from (4), (5), (6), (7) by using the relations

$$(8) \quad \phi_{\alpha\beta\gamma}(x, y + v) = \phi_{\alpha\beta\delta}(x, y + v) \frac{\vartheta_\delta(y + v)}{\vartheta_\gamma(y + v)},$$

$$(9) \quad \phi_{\alpha\beta\gamma}(x, y + v) = \phi_{\alpha\delta\gamma}(x, y + v) \frac{\vartheta_\delta(x)}{\vartheta_\beta(x)}.$$

For example

$$\begin{aligned} (10) \quad (323) &= (331) \frac{\vartheta_1(y + v) \vartheta_3(x)}{\vartheta_3(y + v) \vartheta_2(x)} \\ &= K(311, 122) \{ (001, 332) - (332, 001) \} \frac{\vartheta_3(x)}{\vartheta_2(x)} \\ &= K(311, 122) \{ (001, 322) - (322, 001) \}, \end{aligned}$$

and

$$(11) \quad (010) = K(011, 122) \{ (331, 012) - (012, 331) \}.$$

3. *The addition formulae.* It follows readily from (4), (10), (11) that

$$\begin{aligned} (12) \quad \phi_{331}(x + u, y + v) &= K(111, 122) K'(011, 122) K'(311, 122) \\ &\quad \cdot \{ \phi_{010}(v, x + u) \phi_{323}(y, x + u) - \phi_{323}(v, x + u) \phi_{010}(y, x + u) \} \\ &= K(111, 122) K'(011, 122) K'(311, 122) \\ &\quad \cdot \{ \phi_{322}(x, y) \phi_{021}(x, v) \phi_{010}(u, y) \phi_{313}(u, v) \\ &\quad - \phi_{010}(x, y) \phi_{021}(x, v) \phi_{322}(u, y) \phi_{313}(u, v) \\ &\quad + \phi_{010}(x, y) \phi_{313}(x, v) \phi_{322}(u, y) \phi_{021}(u, v) \\ &\quad - \phi_{322}(x, y) \phi_{313}(x, v) \phi_{010}(u, y) \phi_{021}(u, v) \\ &\quad + \phi_{313}(x, y) \phi_{322}(x, v) \phi_{021}(u, y) \phi_{010}(u, v) \\ &\quad - \phi_{021}(x, y) \phi_{322}(x, v) \phi_{313}(u, y) \phi_{010}(u, v) \\ &\quad + \phi_{021}(x, y) \phi_{010}(x, v) \phi_{313}(u, y) \phi_{322}(u, v) \\ &\quad - \phi_{313}(x, y) \phi_{010}(x, v) \phi_{021}(u, y) \phi_{322}(u, v) \}, \end{aligned}$$

where K' is the same as K with y and v replaced by x and u respectively. Notice that since $\phi_{\alpha\beta\gamma}(x+u, y+v)$ is equal to $\phi_{\alpha\gamma\beta}(y+v, x+u)$ we can obtain a formula for $\phi_{313}(x+u, y+v)$ by interchanging x and y and u and v .

The formulae for all sixty-four functions can be found as above. By increasing the variables in turn by $\pi/2$ and $\pi\tau/2$ we can also obtain other formulae for each function.

In § 2 we used a formula of Jacobi's containing the constant factor ϑ_2 and consequently K and K' both contain ϑ_2 . If we start with a formula containing ϑ_0 or ϑ_3 we get new sets of addition formulae in which the terms corresponding to K and K' contain ϑ_0 or ϑ_3 respectively.

THE UNIVERSITY OF BRITISH COLUMBIA,
VANCOUVER, CANADA.

A THIRD-ORDER IRREGULAR BOUNDARY VALUE PROBLEM AND THE ASSOCIATED SERIES.*

By LEWIS E. WARD.

Introduction. The objects of this paper are to discuss the characteristic functions defined by the system consisting of the differential equation

$$(1) \quad d^3u/dx^3 + [\rho^3 + r(x)]u = 0$$

and the boundary conditions

$$\begin{aligned} W_1(u) &\equiv \alpha_{12}u''(0) + \alpha_{11}u'(0) + \alpha_{10}u(0) = 0, \\ (2) \quad W_2(u) &\equiv \alpha_{22}u''(0) + \alpha_{21}u'(0) + \alpha_{20}u(0) \\ &\quad + \beta_{22}u''(\pi) + \beta_{21}u'(\pi) + \beta_{20}u(\pi) = 0, \\ W_3(u) &\equiv \alpha_{31}u'(0) + \alpha_{30}u(0) = 0, \end{aligned}$$

and to consider the expansion of arbitrary functions in infinite series of these characteristic functions.

In previous papers † on this type of boundary value problem it has been assumed either that the function $r(x)$ appearing in the differential equation possesses a Maclaurin's development in powers of x^3 and that the α 's and β 's are specially chosen, or that $r(x) \equiv 0$ and the α 's and β 's are arbitrary except that a certain determinant of the α 's should not vanish. As a consequence of these assumptions it was found that an arbitrary function which is to be expanded in an infinite series of the characteristic functions must be analytic at $x = 0$ and its Maclaurin's expansion must have a special form.

In the present paper we make no restriction on the form of the function $r(x)$, supposing only that it is continuous in the interval $0 \leq x \leq \pi$ (and for certain theorems either that $r(x)$ has derivatives of all orders on some interval of which $x = 0$ is an interior point, or even that $r(x)$ is analytic at $x = 0$). The hypothesis imposed on the α 's and β 's in a previous paper is retained, that is, they shall be real constants such that the determinant D_α of the α 's arranged as in equations (2) does not vanish, that the matrix

$$\begin{pmatrix} \alpha_{12} & \alpha_{11} & \alpha_{10} \\ 0 & \alpha_{31} & \alpha_{30} \end{pmatrix}$$

* Presented to the American Mathematical Society, February 25, 1933.

† D. Jackson and J. W. Hopkins, *Transactions of the American Mathematical Society*, vol. 20 (1919), p. 245, *et seq.*, and L. E. Ward, *Transactions of the American Mathematical Society*, vol. 29 (1927), p. 716, *et seq.*, and vol. 34 (1932), p. 417, *et seq.*

is of rank two, and that not all the β 's are zero. The removal of restrictions on the function $r(x)$ allows us to offer a proof of the validity of the formal expansion of certain functions not necessarily analytic at $x=0$. Due to this feature the proof has to follow lines somewhat different from those employed previously in irregular boundary value problems.

PART I.

This part of the paper is devoted to a study of the characteristic functions. We first define the three functions *

$$\begin{aligned}\delta_1(t) &= e^{\omega_1 t} + e^{\omega_2 t} + e^{\omega_3 t}, \\ \delta_2(t) &= e^{\omega_1 t} - \omega_3 e^{\omega_2 t} - \omega_2 e^{\omega_3 t}, \\ \delta_3(t) &= e^{\omega_1 t} - \omega_2 e^{\omega_2 t} - \omega_3 e^{\omega_3 t},\end{aligned}$$

in which $\omega_1 = -1$, $\omega_2 = e^{\pi i/3}$, and $\omega_3 = e^{-\pi i/3}$.

THEOREM I. *A necessary and sufficient condition that $u(x, \rho)$ satisfy equation (1) and the first and third of equations (2) is that*

$$\begin{aligned}u(x, \rho) &= k[\alpha_{12}\alpha_{31}\rho^2\delta_1(\rho x) + \alpha_{12}\alpha_{30}\rho\delta_2(\rho x) + (\alpha_{11}\alpha_{30} - \alpha_{10}\alpha_{31})\delta_3(\rho x)] \\ &\quad - (1/3\rho^2) \int_0^x r(t)\delta_3[\rho(x-t)]u(t, \rho)dt,\end{aligned}$$

where k is independent of x .†

To prove the sufficiency we differentiate with respect to x three times both sides of the integral equation in the statement of the theorem. This is seen to result in equation (1). At the same time we verify that the first and third of equations (2) are satisfied.

To prove the necessity we will show that if $u(x, \rho)$ satisfies equation (1), the first and third of equations (2), and also $\alpha_2 u''(0) + \alpha_1 u'(0) + \alpha_0 u(0) = l$, where $l \neq 0$ is given, and $\alpha_2, \alpha_1, \alpha_0$ are chosen so that the determinant

$$D = \begin{vmatrix} \alpha_{12} & \alpha_{11} & \alpha_{10} \\ 0 & \alpha_{31} & \alpha_{30} \\ \alpha_2 & \alpha_1 & \alpha_0 \end{vmatrix}$$

does not vanish, then a value of k , independent of x , exists such that $u(x, \rho)$

* These functions were studied by L. Olivier, *Crelle*, Bd. 2, p. 243. Some of their properties will be found in my 1927 paper, p. 720, already referred to.

† We are concerned only with the solution of equation (1) which is continuous at $x=0$, or if $r(x)$ is analytic at $x=0$, with the solution which is analytic at this point.

In *Comptes Rendus*, t. 90 (1880), p. 721, Y. Villarceau gives the solution of the equation $u^{(m)} \mp rm u = V(x)$. The integral equation of this theorem may be regarded as a special case of Villarceau's formula.

satisfies the integral equation. First we note that $\alpha_2, \alpha_1, \alpha_0$ can be found such that D does not vanish. Hence a unique $u(x, \rho)$ is determined, which depends upon l . On choosing $l = l/(3D\rho^2)$, it is easy to see that the unique solution $\bar{u}(x, \rho)$ of the integral equation satisfies equation (1), the first and third of equations (2), and $\alpha_2 u''(0) + \alpha_1 u'(0) + \alpha_0 u(0) = l$. Hence $\bar{u}(x, \rho) \equiv u(x, \rho)$.

Because of the homogeneous character of equations (1) and (2) we take $k = 1$ without any loss of generality. Instead of obtaining properties of $u(x, \rho)$ from the above integral equation it is desirable to obtain properties of the solution of

$$(3) \quad u(x, \xi, \rho) = U(x, \xi, \rho) - (1/3\rho^2) \int_{\xi}^x r(t) \delta_3[\rho(x-t)] u(t, \xi, \rho) dt,$$

where $U(x, \xi, \rho) \equiv \alpha_{12}\alpha_{31}\rho^2\delta_1[\rho(x-\xi)]$

$$+ \alpha_{12}\alpha_{30}\rho\delta_2[\rho(x-\xi)] + (\alpha_{11}\alpha_{30} - \alpha_{10}\alpha_{31})\delta_3[\rho(x-\xi)],$$

since the function defined by this integral equation enters in a later part of the paper. We note that $u(x, \rho) \equiv u(x, 0, \rho)$.

Let m be the exponent of the highest power of ρ with non-zero coefficient in $U(x, \xi, \rho)$, and denote by S_1 the sector of the ρ -plane defined by $0 \leq \arg \rho \leq \pi/3$. We prove

THEOREM II. *If $0 \leq \xi \leq x \leq \pi$, and if ρ is in S_1 with $|\rho|$ large, then*

$$\begin{aligned} u(x, \xi, \rho) &= U(x, \xi, \rho) + e^{\omega_{3\rho}(x-\xi)} \rho^{m-2} E(x, \xi, \rho),^* \\ u_x'(x, \xi, \rho) &= U_x'(x, \xi, \rho) + e^{\omega_{3\rho}(x-\xi)} \rho^{m-1} E(x, \xi, \rho), \\ u_x''(x, \xi, \rho) &= U_x''(x, \xi, \rho) + e^{\omega_{3\rho}(x-\xi)} \rho^m E(x, \xi, \rho). \end{aligned}$$

If we define $z(x, \xi, \rho)$ by the equation

$$u(x, \xi, \rho) = U(x, \xi, \rho) + e^{\omega_{3\rho}(x-\xi)} z(x, \xi, \rho),$$

we find that $z(x, \xi, \rho)$ satisfies the equation

$$\begin{aligned} z(x, \xi, \rho) &= - (1/3\rho^2) e^{\omega_{3\rho}(\xi-x)} \int_{\xi}^x r(t) \delta_3[\rho(x-t)] U(t, \xi, \rho) dt \\ &\quad - (1/3\rho^2) \int_{\xi}^x r(t) \delta_3[\rho(x-t)] e^{\omega_{3\rho}(t-x)} z(t, \xi, \rho) dt. \end{aligned}$$

If M denotes the maximum of $|z(x, \xi, \rho)|$ for $0 \leq \xi \leq x \leq \pi$, we have, for the values of x and ξ which give $|z(x, \xi, \rho)|$ this maximum

* Throughout this paper we denote by E a function of the indicated variables which is bounded when $|\rho|$ is large. Consequently many different bounded functions will be denoted by the same symbol, but no confusion will arise.

$$M \leq R |3\rho^2|^{-1} |e^{\omega_3\rho(\xi-x)}| \int_{\xi}^x |\delta_3[\rho(x-t)]U(t, \xi, \rho)| dt \\ + RM |3\rho^2|^{-1} \int_{\xi}^x |\delta_3[\rho(x-t)]e^{\omega_3\rho(t-x)}| dt,$$

where $R = \max |r(t)|$ on the interval $0 \leq t \leq \pi$.

But on S_1 we have $|\delta_n[\rho(x-t)]| \leq 3 |e^{\omega_{3n}\rho(x-t)}|$, $n = 1, 2, 3$, and $|U(t, \xi, \rho)| \leq A |\rho^m e^{\omega_{3n}\rho(t-\xi)}|$, where A is independent of t , ξ , and ρ . Also $|\delta_3[\rho(x-t)]e^{\omega_{3n}\rho(t-x)}| \leq 3$. Hence $M \leq RA\pi |\rho|^{m-2} + RM\pi |\rho|^{-2}$. Hence $M \leq B |\rho|^{m-2}$, where B is independent of x , ξ , and ρ . Hence $z(x, \xi, \rho) = \rho^{m-2}E(x, \xi, \rho)$. This gives the first conclusion stated in the theorem.

$$\text{Now } u'_x(x, \xi, \rho) = U'_x(x, \xi, \rho) + (1/3\rho) \int_{\xi}^x r(t)\delta_2[\rho(x-t)]u(t, \xi, \rho) dt.$$

$$\text{Hence } |u'_x(x, \xi, \rho) - U'_x(x, \xi, \rho)| \leq R |3\rho|^{-1} \int_{\xi}^x |\delta_2[\rho(x-t)]u(t, \xi, \rho)| dt.$$

On putting into this integrand the expression found above for $u(t, \xi, \rho)$ and using inequalities similar to those above, we obtain

$$|u'_x(x, \xi, \rho) - U'_x(x, \xi, \rho)| \leq C |\rho^{m-1}e^{\omega_{3n}\rho(x-\xi)}|,$$

where C is independent of x , ξ , ρ , and from this follows the second conclusion stated in the theorem. The final conclusion is obtained in the same way.

The function $u(x, \xi, \rho)$ is analytic in ρ for every finite ρ , and real when x , ξ , and ρ are real. Hence its Maclaurin's expansion in ρ has real coefficients. Hence, denoting conjugates by dashes, $u(x, \xi, \bar{\rho}) = \overline{u(x, \xi, \rho)}$. This fact will be used in the discussion of the characteristic numbers, and also in the third part of the paper.

The characteristic equation. The characteristic equation is $\Delta(\rho) = 0$, where

$$\Delta(\rho) = \begin{vmatrix} W_1(u_1) & W_1(u_2) & W_1(u_3) \\ W_2(u_1) & W_2(u_2) & W_2(u_3) \\ W_3(u_1) & W_3(u_2) & W_3(u_3) \end{vmatrix},$$

and $u_1(x, \rho)$, $u_2(x, \rho)$, $u_3(x, \rho)$ are any three independent solutions of equation (1). We define $u_i(x, \xi, \rho)$, $i = 1, 2, 3$ by

$$(4) \quad u_i(x, \xi, \rho) = \delta_i[\rho(x-\xi)] - (1/3\rho^2) \int_{\xi}^x r(t)\delta_3[\rho(x-t)]u_i(t, \xi, \rho) dt, \\ (i = 1, 2, 3).$$

Evidently these three functions, as functions of x , are solutions of equation (1), and

$$u(x, \xi, \rho) = \alpha_{12}\alpha_{31}\rho^2 u_1(x, \xi, \rho) \\ + \alpha_{12}\alpha_{30}\rho u_2(x, \xi, \rho) + (\alpha_{11}\alpha_{30} - \alpha_{10}\alpha_{31})u_3(x, \xi, \rho).$$

We take $u_i(x, \rho) \equiv u_i(x, 0, \rho)$. Then

$$\begin{aligned} u_1(0, \rho) &= 3 & u_2(0, \rho) &= 0 & u_3(0, \rho) &= 0 \\ u_1'(0, \rho) &= 0 & u_2'(0, \rho) &= -3\rho & u_3'(0, \rho) &= 0 \\ u_1''(0, \rho) &= 0 & u_2''(0, \rho) &= 0 & u_3''(0, \rho) &= 3\rho^2, \end{aligned}$$

and

$$\Delta(\rho) = \begin{vmatrix} 3\alpha_{10} & -3\alpha_{11}\rho & 3\alpha_{12}\rho^2 \\ 3\alpha_{20} + W_{2\pi}(u_1) & -3\alpha_{21}\rho + W_{2\pi}(u_2) & 3\alpha_{22}\rho^2 + W_{2\pi}(u_3) \\ 3\alpha_{30} & -3\alpha_{31}\rho & 0 \end{vmatrix},$$

where $W_{2\pi}(u_i) \equiv \beta_{22}u_i''(\pi, \rho) + \beta_{21}u_i'(\pi, \rho) + \beta_{20}u_i(\pi, \rho)$.

On expanding the determinant for $\Delta(\rho)$ we obtain

$$\Delta(\rho) = 27D_a\rho^3 - 9\rho W_{2\pi}(u).$$

If we let β_{2j} be that β not equal to zero with the highest second subscript, and use the expressions given in Theorem II for $u(x, \rho)$ and its derivatives, we have

$$\Delta(\rho) = 27D_a\rho^3 + \rho^{m+j+1}e^{\omega_3\rho\pi}[A\delta_k(\rho\pi)e^{-\omega_3\rho\pi} + \rho^{-1}E(\rho)],$$

where A is independent of ρ and is not zero, and k is one of the numbers 1, 2, 3. Hence

$$\Delta(\rho) = \rho^{m+j+1}e^{\omega_3\rho\pi}[A\delta_k(\rho\pi)e^{-\omega_3\rho\pi} + \rho^{-1}E(\rho)].$$

This form is valid if ρ is in the sector S_1 and $|\rho|$ is large.

For $|\rho|$ large the function $\delta_k(\rho\pi)e^{-\omega_3\rho\pi}$ is known to have zeros ρ'_n which are simple and real, with successive zeros separated from one another by a distance which is uniformly bounded from zero. Furthermore, if we construct small circles all of the same radius, centered at the points ρ'_n , and call S'_1 the part of S_1 not inside these circles, we have in S'_1 $|\delta_k(\rho\pi)e^{-\omega_3\rho\pi}| > \delta$, where δ is independent of ρ and is positive.* Hence for $|\rho|$ sufficiently large and ρ in S'_1 we have

$$(5) \quad |\Delta(\rho)| > h |\rho^{m+j+1}e^{\omega_3\rho\pi}|,$$

where h is independent of ρ .

We denote by S_2 and S'_2 the reflections of S_1 and S'_1 in the axis of reals. Then, since $\Delta(\rho)$ takes on in S_2 values conjugate to those it has in S_1 , we

* Ward, *loc. cit.*, 1927, pp. 718 and 719.

have in S'_2 for $|\rho|$ large $\Delta(\rho) > h |\rho^{m+j+1} e^{\omega_2 \rho \pi}|$. Hence for $|\rho|$ large the zeros of $\Delta(\rho)$ can occur only in the small circles. That there is just one in each such small circle and that it is real is shown in the usual way.* These zeros are the characteristic numbers, and are denoted in succession by ρ_1, ρ_2, \dots .

The characteristic functions. The $U(x, \xi, \rho)$ of Theorem II is identical with the $u(x)$ in equation (4) on page 720 of the 1927 paper if a is replaced by ξ . Hence by formula (5) of that paper

$$\begin{aligned} u(x, \xi, \rho) = & e^{-\rho(x-\xi)} [\alpha_{12}\alpha_{31}\rho^2 + \alpha_{12}\alpha_{30}\rho + (\alpha_{11}\alpha_{30} - \alpha_{10}\alpha_{31})] \\ & + 2e^{\rho(x-\xi)/2} [\alpha_{12}\alpha_{31}\rho^2 \cos \{3^{1/2}\rho(x-\xi)/2\} \\ & - \alpha_{12}\alpha_{30}\rho \cos \{-\pi/3 + 3^{1/2}\rho(x-\xi)/2\} \\ & - (\alpha_{11}\alpha_{30} - \alpha_{10}\alpha_{31}) \cos \{\pi/3 + 3^{1/2}\rho(x-\xi)/2\}] \\ & + e^{\omega_2 \rho(x-\xi)} \rho^{m-2} F(x, \xi, \rho). \end{aligned}$$

On putting $\xi = 0$, and $\rho = \rho_k$, we obtain the characteristic functions of the present paper in the form

$$(6) \quad u_k(x) = 2e^{\rho_k x/2} \left[\begin{aligned} & \alpha_{12}\alpha_{31}\rho_k^2 \cos(3^{1/2}\rho_k x/2) - \alpha_{12}\alpha_{30}\rho_k \cos(-\pi/3 + 3^{1/2}\rho_k x/2) \\ & - (\alpha_{11}\alpha_{30} - \alpha_{10}\alpha_{31}) \cos(\pi/3 + 3^{1/2}\rho_k x/2) \\ & + e^{-3\rho_k x/2} \{ \alpha_{12}\alpha_{31}\rho_k^2 + \alpha_{12}\alpha_{30}\rho_k + (\alpha_{11}\alpha_{30} - \alpha_{10}\alpha_{31}) \} / 2 \\ & + \rho_k^{m-2} F(x, \rho_k) \end{aligned} \right]$$

Since ρ_k is real, at least when k is sufficiently large, this form shows clearly the dominant terms in $u_k(x)$.

PART II.

We consider now infinite series of the above characteristic functions,

$$(7) \quad \sum_{k=1}^{\infty} a_k u_k(x),$$

where the a 's are independent of x , and we shall derive certain properties of the sum of such a series. We prove first

THEOREM III. *If series (7) converges uniformly in $0 \leq \alpha \leq x \leq x_0 \leq \pi$, where $\alpha < x_0$, and x_1 is any number less than x_0 , then $|a_k| < \gamma \rho_k^{-m} e^{-\rho_k x_1/2}$, where γ is independent of k .*

If k is sufficiently large, we can find a number x'_k in (x_1, x_0) such that any one of the cosines in equation (6) has the value unity for $x = x'_k$. Hence

* Ward, *loc. cit.*, 1932, p. 420.

$|u_k(x'_k)| > \gamma' \rho_k^m e^{\rho_k x'_k/2}$, where γ' is independent of k . But $|a_k u_k(x)| < \gamma''$, where γ'' is independent of x and of k . Hence $|a_k| < \gamma \rho_k^{-m} e^{-\rho_k x'_k/2} < \gamma \rho_k^{-m} e^{-\rho_k x_1/2}$. This inequality can be extended to include all values of k by choosing a different γ if necessary.

THEOREM IV. *If $r(x)$ has derivatives of all orders on the interval $-x_0/2 \leq x \leq x_0$, and if the hypothesis of Theorem III is satisfied, then the sum $f(x)$ of series (7) possesses continuous derivatives of all orders in the interval $-x_2/2 \leq x \leq x_2$, where $0 < x_2 < x_1$.*

It is clear from equation (3) and the equations obtained from it by successive differentiations with respect to x that, since $r(x)$ has derivatives of all orders in the interval $-x_0/2 \leq x \leq x_0$, the functions $u_k(x)$ will also have derivatives of all orders on this interval, and these derivatives will all be continuous. Also, successive repetitions with slight variations of the argument of Theorem II show that $|u_k^{(j)}(x)| < L_j \rho_k^{m+j} e^{\rho_k x/2}$ if $x \geq 0$, and $|u_k^{(j)}(x)| < L_j \rho_k^{m+j} e^{-\rho_k x}$ if $x \leq 0$, where L_j is independent of k and of x . Hence, if x is in the interval $-x_2/2 \leq x \leq x_2$, we have $|a_k u_k^{(j)}(x)| < \gamma L_j \rho_k^j e^{\rho_k(x_2-x_1)/2}$. But for each j this is the general term of a convergent series of positive constants, and the series $\sum a_k u_k^{(j)}(x)$ converges uniformly in the interval $-x_2/2 \leq x \leq x_2$, j being any positive integer or zero. From this follows the conclusion stated in the theorem.

Let us define the w 's by means of the equations

$$w_0(x) = f(x), \quad w_n(x) = w''_{n-1}(x) + r(x)w_{n-1}(x), \quad (n = 1, 2, 3, \dots).$$

Then $w_n(x) = (-1)^n \sum_{k=1}^{\infty} a_k \rho_k^{3n} u_k(x)$. Hence by the first and third of equations (2)

$$(8) \quad \left. \begin{aligned} \alpha_{12} w''_n(0) + \alpha_{11} w'_n(0) + \alpha_{10} w_n(0) &= 0 \\ \alpha_{31} w'_n(0) + \alpha_{30} w_n(0) &= 0 \end{aligned} \right\} \quad (n = 0, 1, 2, \dots).$$

We have, therefore,

THEOREM V. *If $r(x)$ has derivatives of all orders in an interval of which $x = 0$ is an interior point, and if the hypothesis of Theorem III is satisfied, then the sum $f(x)$ of series (7) possesses derivatives of all orders at $x = 0$, which satisfy the infinite set of equations (8).*

Equations (8) consist of an infinite set of linear homogeneous equations connecting the values of the derivatives of $f(x)$ at $x = 0$. If they be grouped in pairs, the first pair arising from $n = 0$, the second from $n = 1$, etc., it is

evident from the first pair that one of $f(0)$, $f'(0)$, $f''(0)$ can be chosen arbitrarily, from the second pair that the corresponding one of $f'''(0)$, $f^{IV}(0)$, $f^V(0)$ can be chosen arbitrarily, and so on. The remaining derivatives then have unique values.

This indicates the degree of arbitrariness in $f(x)$. However, some further restriction beyond equations (8) must be made in order to establish the convergence to $f(x)$ of the formal series. The particular restriction made in this paper is not a necessary condition on $f(x)$, and its statement will be postponed to Part III.

In order to discuss the convergence of series (7) for complex values of x it is desirable to have the asymptotic forms of $u_k(x)$ for large k and for x in certain regions to be defined presently. In order to obtain these forms we shall use equation (3) with $\xi = 0$, allowing x to be a complex variable and ρ a positive constant, and we shall suppose $r(x)$ to be analytic at $x = 0$. We shall take the t -integration over a single straight line. The existence of a unique solution of (3) analytic in x provided x is inside the region containing $x = 0$ in which $r(x)$ is analytic can be shown in the usual way.* We now prove

THEOREM VI. *If $r(x)$ is analytic at $x = 0$ and if T_3 is the finite part of the sector $0 \leq \arg x \leq 2\pi/3$, including the boundaries, cut off by a straight line drawn so that T_3 contains no singularity of $r(x)$, then in T_3 we have $u(x, \rho) = U(x, 0, \rho) + e^{\omega_3 \rho x} \rho^{m-2} E(x, \rho)$, where $E(x, \rho)$ is bounded and analytic in x for ρ large and positive. If T_2 and T_1 are regions similarly constructed in the sectors $4\pi/3 \leq \arg x \leq 2\pi$ and $2\pi/3 \leq \arg x \leq 4\pi/3$ respectively, then*

$$\begin{aligned} u(x, \rho) &= U(x, 0, \rho) + e^{\omega_2 \rho x} \rho^{m-2} E(x, \rho) \text{ in } T_2, \text{ and} \\ u(x, \rho) &= U(x, 0, \rho) + e^{\omega_1 \rho x} \rho^{m-2} E(x, \rho) \text{ in } T_1. \end{aligned}$$

To give the proof for the region T_3 we write $u(x, \rho) = U(x, 0, \rho) + e^{\omega_3 \rho x} \rho^{m-2} z(x, \rho)$. From equation (3) we see that $z(x, \rho)$ will satisfy the integral equation

$$\begin{aligned} z(x, \rho) &= -\rho^{-m} \int_0^x r(t) \delta_3[\rho(x-t)] e^{-\omega_3 \rho x} U(t, 0, \rho) dt \\ &\quad - (1/3\rho^2) \int_0^x r(t) \delta_3[\rho(x-t)] e^{\omega_3 \rho(t-x)} z(t, \rho) dt. \end{aligned}$$

From its definition it is clear that $z(x, \rho)$ is an analytic function of x in the closed region T_3 . Let $|z(x, \rho)|$ attain its maximum M in T_3 for $x = x_3$. Then

* See the 1932 paper, pp. 421 and 422, where the proof is given for a special case.

for $x = x_3$ we have $M \leq |E_1(\rho)| + M |E_2(\rho)\rho^{-2}|$, whence M is a bounded function of ρ , and $z(x, \rho)$ is a bounded function of x and of ρ .

The proofs for the regions T_2 and T_1 are given in a similar way.

We can now consider the convergence of series (7) for complex values of x . Let T_3 , T_2 , and T_1 be such that they form an equilateral triangle T_{x_2} whose center is at $x = 0$ and one vertex of which is at the point $x = x_2$ on the positive axis of reals.* By Theorem VI we have in T_{x_2} $|u(x, \rho)| \leq c\rho^m e^{\rho x_2/2}$, where c is independent of x and of ρ . If we suppose the hypothesis of Theorem III is satisfied, then $|a_k u_k(x)| < c\gamma e^{\rho k(x_2 - x_1)/2}$. If we now take $0 < x_2 < x_1$; the last expression is the general term of a convergent series of positive constants, and series (7) converges uniformly in the interior and on the boundary of T_{x_2} . We have, therefore,

THEOREM VII. *If $r(x)$ is analytic at $x = 0$ and if the hypothesis of Theorem III is satisfied, then series (7) converges uniformly in the interior and on the boundary of an equilateral triangle T_{x_2} centered at $x = 0$ and having one vertex at $x = x_2$ on the axis of reals between $x = 0$ and $x = x_0$, provided T_{x_2} does not have in its interior or on its boundary a singularity of $r(x)$.*

THEOREM VIII. *If X is the upper limit of all possible choices of the x_0 of Theorem III, if $y > X$, and if $r(x)$ has no singularity inside T_y , then series (7) cannot converge at any point outside T_X but inside T_y except possibly points on the rays $\arg x = 0, 2\pi/3, 4\pi/3$.*

We omit the proof, which follows the same lines as the proof of Theorem VII, page 423 of the 1932 paper.†

The derivation of equations (8) satisfied by the analytic sum $f(x)$ of series (7) is the same as in the case where the mere existence of all derivatives of $f(x)$ and of $r(x)$ was known. Accordingly we have

THEOREM IX. *If $r(x)$ is analytic at $x = 0$ and if the hypothesis of Theorem III is satisfied, then series (7) converges to a function $f(x)$ analytic at $x = 0$ and satisfying equations (8).*

* By the notation T_a we shall mean an equilateral triangle centered at $x = 0$ with one vertex at $x = a$, $a > 0$.

† In the proof there given the point x'_2 is supposed to be such that $0 < \arg x'_2 < 2\pi/3$ instead of $0 \leq \arg x'_2 \leq 2\pi/3$, as was incorrectly stated.

PART III.

By the formal series for $f(x)$ we mean a series of type (7) in which the a 's are determined by means of certain orthogonality relations involving the adjoint characteristic functions.* It is known that the sum of the first n terms of the formal series for $f(x)$ equals the contour integral

$$(1/2\pi i) \int_{\gamma_n} \int_0^\pi 3\rho^2 f(t) G(x, t, \rho) dt d\rho, \dagger$$

where $G(x, t, \rho)$ is the Green's function of the system (1) and (2), and γ_n is the arc of a circle centered at $\rho = 0$, of radius between ρ_n and ρ_{n+1} , and extending from the ray $\arg \rho = -\pi/3$ to the ray $\arg \rho = \pi/3$.

A formula for $G(x, t, \rho)$ useful in the present case is given on page 723 of the 1927 paper. The function $g(x, t, \rho)$ there defined is given by

$$g(x, t, \rho) = \pm (1/2) \sum_{j=1}^3 u_j(x) v_j(t), \quad + \text{ if } x > t, \quad - \text{ if } x < t,$$

where the u 's are any three independent solutions of equation (1), and $v_j(t)$ is the cofactor of $u''_j(t)$ in the determinant

$$W = \begin{vmatrix} u''_1(t) & u''_2(t) & u''_3(t) \\ u'_1(t) & u'_2(t) & u'_3(t) \\ u_1(t) & u_2(t) & u_3(t) \end{vmatrix} \text{ divided by } W.$$

It is easy to show that the function $\phi(x) = 3\rho^2 \sum_{j=1}^3 u_j(x) v_j(\xi)$ satisfies the integral equation (4) with $i = 3$. Hence

$$g(x, t, \rho) = \pm u_3(x, t, \rho) / (6\rho^2), \quad + \text{ if } x > t, \quad - \text{ if } x < t.$$

The formula for $G(x, t, \rho)$ is $G(x, t, \rho) = -N(x, t, \rho) / \Delta(\rho)$, where

$$N(x, t, \rho) = \begin{vmatrix} u_1(x) & u_2(x) & u_3(x) & g(x, t, \rho) \\ W_1(u_1) & W_1(u_2) & W_1(u_3) & W_1(g) \\ W_2(u_1) & W_2(u_2) & W_2(u_3) & W_2(g) \\ W_3(u_1) & W_3(u_2) & W_3(u_3) & W_3(g) \end{vmatrix}.$$

We note that $\Delta(\rho)$ is the minor of $g(x, t, \rho)$ in $N(x, t, \rho)$.

The Green's function is independent of the manner in which $u_1(x)$, $u_2(x)$,

* See the fundamental paper by Birkhoff, *Transactions of the American Mathematical Society*, vol. 9 (1903), p. 373, et seq.

† Birkhoff, *loc. cit.*, p. 379.

and $u_3(x)$ be chosen, so long as they are independent solutions of equation (1). We shall take for them the functions defined by equation (4) for $\xi = 0$. This gives

$$N(x, t, \rho) = \begin{vmatrix} u_1(x) & u_2(x) & u_3(x) & g(x, t, \rho) \\ 3\alpha_{10} & -3\alpha_{11}\rho & 3\alpha_{12}\rho^2 & W_1(g) \\ 3\alpha_{20} + W_{2\pi}(u_1) & -3\alpha_{21}\rho + W_{2\pi}(u_2) & 3\alpha_{22}\rho^2 + W_{2\pi}(u_3) & W_2(g) \\ 3\alpha_{30} & -3\alpha_{31}\rho & 0 & W_3(g) \end{vmatrix}.$$

In order to evaluate this determinant we multiply the elements in the first three columns by $v_1(t)/2$, $v_2(t)/2$, and $v_3(t)/2$ respectively, and add these products to the elements in the fourth column. This gives zeros for the second and fourth elements of the fourth column. On expanding by minors of the elements of the fourth column we obtain

$$N(x, t, \rho) = -\Delta(\rho) [g(x, t, \rho) + u_3(x, t, \rho)/(6\rho^2)] - 18\rho u(x) W_{2\pi}(g).$$

But $W_{2\pi}(g) = W_{2\pi}(u_3)/(6\rho^2)$. Hence

$$\begin{aligned} G(x, t, \rho) &= u_3(x, t, \rho)/(3\rho^2) + 3u(x) W_{2\pi}(u_3)/[\rho\Delta(\rho)] & \text{if } x > t, \\ &= 3u(x) W_{2\pi}(u_3)/[\rho\Delta(\rho)] & \text{if } x < t. \end{aligned}$$

Denoting by $I_n(x)$ the sum of the first n terms of the formal series for $f(x)$, we now have

$$\begin{aligned} I_n(x) &= \frac{1}{2\pi i} \int_{\gamma_n} \int_0^x f(t) u_3(x, t, \rho) dt d\rho \\ &\quad + \frac{1}{2\pi i} \int_{\gamma_n} \frac{9\rho u(x)}{\Delta(\rho)} \int_0^\pi f(t) W_{2\pi}(u_3) dt d\rho. \end{aligned}$$

We introduce the function $\sigma(x, s) = \int_0^x f(t) u_3(s, t, \rho) dt$, which will be useful in transforming the integrands of the ρ -integrals in $I_n(x)$. Concerning this function we have first the following theorem.

THEOREM X. *The function $\sigma(x, s)$ satisfies the integral equation*

$$\begin{aligned} (9) \quad \sigma(x, s) &= \int_0^x f(t) \delta_3[\rho(s-t)] dt - \frac{1}{3\rho^2} \int_0^x r(t) \delta_3[\rho(s-t)] \sigma(t, t) dt \\ &\quad - \frac{1}{3\rho^2} \int_x^s r(t) \delta_3[\rho(s-t)] \sigma(x, t) dt. \end{aligned}$$

This theorem is a restatement of Theorem X of the 1932 paper.

If we put $s = x$, we obtain from equation (9)

$$(10) \quad \sigma(x) = \int_0^x f(t) \delta_3[\rho(x-t)] dt - \frac{1}{3\rho^2} \int_0^x r(t) \delta_3[\rho(x-t)] \sigma(t) dt,$$

where we have written $\sigma(x) = \sigma(x, x)$.

Before treating the general case it is interesting to consider the special case in which $w_1(x) \equiv 0$. This is the case in which $f(x)$ is a solution of the differential equation $f''' + r(x)f = 0$. We shall suppose that both $f(x)$ and $r(x)$ have derivatives of all orders in the interval $0 \leq x \leq \pi$. On integrating by parts three times the first integral in equation (10), that equation becomes

$$\sigma(x) = 3f(x)/\rho - f(0)\delta_1(\rho x)/\rho + f'(0)\delta_2(\rho x)/\rho^2 - f''(0)\delta_3(\rho x)/\rho^3 \\ - \frac{1}{\rho^3} \int_0^x f'''(t)\delta_3[\rho(x-t)]dt - \frac{1}{3\rho^2} \int_0^x r(t)\delta_3[\rho(x-t)]\sigma(t)dt.$$

Now define $\xi(x)$ by the equation $\sigma(x) = 3f(x)/\rho + \xi(x)$. Then $\xi(x)$ satisfies the integral equation

$$\xi(x) = -f(0)\delta_1(\rho x)/\rho + f'(0)\delta_2(\rho x)/\rho^2 - f''(0)\delta_3(\rho x)/\rho^3 \\ - \frac{1}{3\rho^2} \int_0^x r(t)\delta_3[\rho(x-t)]\xi(t)dt,$$

in the derivation of which we used the fact that $w_1(x) \equiv 0$. But $\alpha_{12}f''(0) + \alpha_{11}f'(0) + \alpha_{10}f(0) = 0$ and $\alpha_{31}f'(0) + \alpha_{30}f(0) = 0$. Hence $f(0) = \lambda\alpha_{12}\alpha_{31}$, $f'(0) = -\lambda\alpha_{12}\alpha_{30}$, $f''(0) = \lambda(\alpha_{11}\alpha_{30} - \alpha_{10}\alpha_{31})$, where λ is a non-vanishing constant independent of ρ . Hence

$$\xi(x) = -\lambda U(x, 0, \rho)/\rho^3 - (1/3\rho^2) \int_0^x r(t)\delta_3[\rho(x-t)]\xi(t)dt.$$

On comparing this equation with equation (3) for $\xi = 0$ we infer that $\xi(x) = -\lambda u(x, 0, \rho)/\rho^3$. Hence we have

$$(11) \quad \sigma(x) = 3f(x)/\rho - \lambda u(x, \rho)/\rho^3.$$

It is in the obtaining of this equation that the necessary conditions (8) enter.

As for the second ρ -integral in $I_n(x)$, we have

$$\int_0^\pi f(t)W_{2\pi}(u_3)dt = \beta_{22} \int_0^\pi f(t)u''_3(\pi, t, \rho)dt + \beta_{21} \int_0^\pi f(t)u'_3(\pi, t, \rho)dt \\ + \beta_{20} \int_0^\pi f(t)u_3(\pi, t, \rho)dt,$$

where the accents mean derivatives with respect to the first indicated argument.

Now from $\sigma(x) = \int_0^x f(t)u_3(x, t, \rho)dt$ we have $\sigma'(x) = \int_0^x f(t)u'_3(x, t, \rho)dt$, since $u_3(x, x, \rho) \equiv 0$. Similarly $\sigma''(x) = \int_0^x f(t)u''_3(x, t, \rho)dt$. Hence $\int_0^\pi f(t)W_{2\pi}(u_3)dt = 3W_{2\pi}(f)/\rho - \lambda W_{2\pi}(u)/\rho^3$.

We have, therefore,

$$\begin{aligned} I_n(x) &= \frac{1}{2\pi i} \int_{\gamma_n} [3f(x)/\rho - \lambda u(x)/\rho^3 \\ &\quad + \frac{u(x)}{3D_a\rho^2 - W_{2\pi}(u)} \{3W_{2\pi}(f)/\rho - \lambda W_{2\pi}(u)/\rho^3\}] d\rho \\ &= \frac{1}{2\pi i} \int_{\gamma_n} [3f(x)/\rho + 2\gamma u(x) \{W_{2\pi}(f) - \lambda D_a\}/\Delta(\rho)] d\rho. \end{aligned}$$

The cancelling of two large terms in this integrand was due to the form of $\sigma(x)$, which goes back to the form of $f(x)$ imposed in accordance with necessary conditions (8).

On account of the conjugate property of $u(x, \rho)$ in ρ , already referred to, we have

$$I_n(x) = \frac{1}{\pi i} \int_{\gamma'_n} [3f(x)/\rho + 2\gamma u(x) \{W_{2\pi}(f) - \lambda D_a\}/\Delta(\rho)] d\rho,$$

where γ'_n is the part of γ_n in S_1 . But in S_1 we have $u(x) = \rho^m e^{i\omega_3 \rho x} E(x, \rho)$, while $W_{2\pi}(f) - \lambda D_a$ is independent of ρ . Recalling inequality (5) we see that $I_n(x) = f(x) + \epsilon_n(x)$, where $\epsilon_n(x)$ tends uniformly to zero as n becomes infinite, x being in the interval $0 \leq x \leq \beta < \pi$, where β is any constant between 0 and π . Consequently the formal series for $f(x)$ converges uniformly to $f(x)$ in the interval $0 \leq x \leq \beta$. A similar discussion of the convergence of the formal series can be given if $w_k(x)$, $k > 1$, vanishes identically.

If $r(x)$ is analytic at $x = 0$, the uniform convergence of the formal series may be extended to appropriate regions of the x -plane by using Theorems VII and VIII. The largest region of uniform convergence may not be an equilateral triangle. Its shape depends upon the locations of the singularities of $f(x)$ and $r(x)$, and is not discussed here.

In the general case no such simple expression for $\sigma(x)$ as that in equation (11) can be obtained. We shall assume $f(x)$ to possess derivatives of all orders in an interval of which $x = 0$ is an interior point and to possess a continuous second derivative in the interval $0 \leq x \leq \pi$. A different form for the integrand of the second ρ -integral in $I_n(x)$ is desirable, and we proceed to the derivation of this. We have

$$\int_0^x f(t) W_{2\pi}(u_3) dt = \int_0^x f(t) W_{2\pi}(u_3) dt + \int_x^\pi f(t) W_{2\pi}(u_3) dt.$$

Transforming the first integral on the right in an obvious way results in

$$\int_0^\pi f(t) W_{2\pi}(u_3) dt = \beta_{22}\sigma''(x, \pi) + \beta_{21}\sigma'(x, \pi) + \beta_{20}\sigma(x, \pi) + \int_x^\pi f(t) W_{2\pi}(u_3) dt,$$

where the accents mean differentiation with respect to s . This gives

$$(12) \quad I_n(x) = \frac{1}{2\pi i} \int_{\gamma_n} \left\{ \sigma(x) + \frac{9\rho u(x)}{\Delta(\rho)} \left\{ \beta_{22}\sigma''(x, \pi) + \beta_{21}\sigma'(x, \pi) + \beta_{20}\sigma(x, \pi) \right\} + \int_x^\pi f(t) W_{2\pi}(u_s) dt \right\} d\rho.$$

We shall now obtain further properties of the function $\sigma(x, s)$, in which we are interested for $x \leq s \leq \pi$. We assume $r(x)$ to possess derivatives of all orders in an interval of which $x = 0$ is an interior point and that the series

$$(13) \quad \frac{3}{\rho} f(x) - \frac{3}{\rho^4} w_1(x) + \frac{3}{\rho^7} w_2(x) - \cdots$$

converges uniformly in some closed interval J of which $x = 0$ is an interior point.* The latter assumption takes the place of the assumption (made in previous papers) that $f(x)$ be analytic at $x = 0$. It could be lightened, but it is made in this form so that we may have a form of solution of equation (10) to which we can apply equations (8) readily.

Using the defining equations of the w 's we see from (13) that the series

$$\frac{3}{\rho} f(x) - \frac{3}{\rho^4} [f'''(x) + r(x)f(x)] + \frac{3}{\rho^7} [w_1'''(x) + r(x)w_1(x)] - \cdots$$

converges uniformly in J , and hence, by subtraction, that

$$\frac{3}{\rho} f'''(x) - \frac{3}{\rho^4} w_1'''(x) + \frac{3}{\rho^7} w_2'''(x) - \cdots$$

converges uniformly in J . Hence, by integration, the series

$$\frac{3}{\rho} [f''(x) - f''(0)] - \frac{3}{\rho^4} [w_1''(x) - w_1''(0)] + \cdots$$

converges uniformly in J . But $\frac{3}{\rho} f''(0) - \frac{3}{\rho^4} w_1''(0) + \cdots$ converges, its terms being proportional to those of (13) at $x = 0$ by equations (8). Hence

$$\frac{3}{\rho} f''(x) - \frac{3}{\rho^4} w_1''(x) + \frac{3}{\rho^7} w_2''(x) - \cdots$$

converges uniformly in J , as does also

$$\frac{3}{\rho} f'(x) - \frac{3}{\rho^4} w_1'(x) + \frac{3}{\rho^7} w_2'(x) - \cdots$$

* The uniform character of the convergence is not necessary for the argument, but is made for convenience.

Consequently, denoting by $\tau(x, \rho)$ the sum of series (13), the x -derivatives of $\tau(x, \rho)$ are obtained by differentiating series (13) termwise.

A set of equations equivalent to (8) is

$$\left. \begin{aligned} w_n(0) &= \lambda_n \alpha_{12} \alpha_{31} \\ w'_n(0) &= -\lambda_n \alpha_{12} \alpha_{30} \\ w''_n(0) &= \lambda_n (\alpha_{11} \alpha_{30} - \alpha_{10} \alpha_{31}) \end{aligned} \right\} \quad (n = 0, 1, 2, \dots).$$

These equations serve to define uniquely the λ 's, which are independent of ρ . Furthermore the series

$$\lambda_0/\rho - \lambda_1/\rho^4 + \lambda_2/\rho^7 - \dots$$

converges, and we denote its sum by $\nu(\rho)$.

THEOREM XI. *If $f(x)$ satisfies equations (8), then $\tau(x, \rho)$ satisfies*

$$\begin{aligned} \tau''' + [\rho^3 + r(x)]\tau &= 3\rho^2 f(x) \\ \tau(0, \rho) &= 3\alpha_{12}\alpha_{31}\nu(\rho) \\ \tau'(0, \rho) &= -3\alpha_{12}\alpha_{30}\nu(\rho) \\ \tau''(0, \rho) &= 3(\alpha_{11}\alpha_{30} - \alpha_{10}\alpha_{31})\nu(\rho). \end{aligned}$$

These are proved immediately by making use of equations (8) and the series for $\tau(x, \rho)$ and its x -derivatives given above.

THEOREM XII. *If $f(x)$ satisfies equations (8), then $\tau(x, \rho)$ satisfies the integral equation*

$$\begin{aligned} \tau(x, \rho) &= \frac{\nu(\rho)}{\rho^2} U(x, 0, \rho) \\ &+ \int_0^x f(t) \delta_3[\rho(x-t)] dt - \frac{1}{3\rho^2} \int_0^x r(t) \delta_3[\rho(x-t)] \tau(t, \rho) dt. \end{aligned}$$

This is an integral equation equivalent to the differential system in the preceding theorem.

The next theorem gives a form for $\sigma(x)$ analogous to that of the special case treated above.

THEOREM XIII. *If $f(x)$ satisfies equations (8), then*

$$\sigma(x) = \tau(x, \rho) - \frac{\nu(\rho)}{\rho^2} u(x).$$

This follows immediately from equation (10), equation (3) with $\xi = 0$, and the equation of Theorem 12. We note that the first term in $\nu(\rho)/\rho^2$, namely, λ_0/ρ^3 , is the negative of the coefficient of $u(x, \rho)$ in equation (11).

THEOREM XIV. If $f(x)$ satisfies equations (8), then $\sigma(x, s)$ satisfies the integral equation

$$\begin{aligned}\sigma(x, s) = & \frac{1}{3\rho^2}[\rho^2\delta_1[\rho(s-x)]\tau(x, \rho) - \rho\delta_2[\rho(s-x)]\tau'(x, \rho) + \delta_3[\rho(s-x)]\tau''(x, \rho)] \\ & - \frac{\nu(\rho)}{\rho^2}u(s, 0, \rho) - \frac{\nu(\rho)}{3\rho^4}\int_x^s r(t)\delta_3[\rho(s-t)]u(t, 0, \rho)dt \\ & - \frac{1}{3\rho^2}\int_x^s r(t)\delta_3[\rho(s-t)]\sigma(x, t)dt.\end{aligned}$$

We insert the expression for $\sigma(x)$ obtained in Theorem 13 into equation (9). This gives

$$\begin{aligned}\sigma(x, s) = & \int_0^x f(t)\delta_3[\rho(s-t)]dt + \frac{\nu(\rho)}{3\rho^4}\int_0^x r(t)\delta_3[\rho(s-t)]u(t, \rho)dt \\ & - \frac{1}{3\rho^2}\int_0^x r(t)\delta_3[\rho(s-t)]\tau(t, \rho)dt - \frac{1}{3\rho^2}\int_x^s r(t)\delta_3[\rho(s-t)]\sigma(x, t)dt.\end{aligned}$$

Using Theorem 11 we have for the third integral in this equation

$$\begin{aligned}& \int_0^x r(t)\delta_3[\rho(s-t)]\tau(t, \rho)dt \\ &= \int_0^x \delta_3[\rho(s-t)][3\rho^2f(t) - \tau'''(t, \rho) - \rho^3\tau(t, \rho)]dt \\ &= 3\rho^2\int_0^x f(t)\delta_3[\rho(s-t)]dt - \rho^3\int_0^x \delta_3[\rho(s-t)]\tau(t, \rho)dt \\ &\quad - \int_0^x \delta_3[\rho(s-t)]\tau'''(t, \rho)dt.\end{aligned}$$

But, integrating by parts three times

$$\begin{aligned}\int_0^x \delta_3[\rho(s-t)]\tau'''(t, \rho)dt = & \delta_3[\rho(s-x)]\tau''(x, \rho) - \rho\delta_2[\rho(s-x)]\tau'(x, \rho) \\ & + \rho^2\delta_1[\rho(s-x)]\tau(x, \rho) - \delta_3(\rho s)\tau''(0, \rho) \\ & + \rho\delta_2(\rho s)\tau'(0, \rho) - \rho^2\delta_1(\rho s)\tau(0, \rho) \\ & - \rho^3\int_0^x \delta_3[\rho(s-t)]\tau(t, \rho)dt.\end{aligned}$$

Hence

$$\begin{aligned}\sigma(x, s) = & [\delta_3[\rho(s-x)]\tau''(x, \rho) - \rho\delta_2[\rho(s-x)]\tau'(x, \rho) \\ & + \rho^2\delta_1[\rho(s-x)]\tau(x, \rho)]/(3\rho^2) \\ & - \frac{\nu(\rho)}{\rho^2}[U(s, 0, \rho) - \frac{1}{3\rho^2}\int_0^x r(t)\delta_3[\rho(s-t)]u(t, \rho)dt] \\ & - \frac{1}{3\rho^2}\int_x^s r(t)\delta_3[\rho(s-t)]\sigma(x, t)dt.\end{aligned}$$

On making use of equation (3) with $x=s$ and $\xi=0$ this becomes the equation of the present theorem.

From the equation of Theorem 14 the desired asymptotic forms of $\sigma(x, s)$ and its s -derivatives can be obtained. Let us write

$$\sigma(x, s) = -v(\rho)u(s, 0, \rho)/\rho^2 + \delta_1[\rho(s-x)]f(x)/\rho + e^{\omega_3\rho(s-x)}v(x, s, \rho)/\rho^2.$$

Then $v(x, s, \rho)$ satisfies the equation

$$\begin{aligned} v(x, s, \rho) = & e^{\omega_3\rho(x-s)}[\rho^2\delta_1[\rho(s-x)]\{\tau(x, \rho) - 3f(x)/\rho\} \\ & - \rho\delta_2[\rho(s-x)]\tau'(x, \rho) + \delta_3[\rho(s-x)]\tau''(x, \rho)]/3 \\ & - \frac{1}{3\rho^2}f(x)e^{\omega_3\rho(x-s)}\int_x^s r(t)\delta_3[\rho(s-t)]\delta_1[\rho(t-x)]dt \\ & - \frac{1}{3\rho^2}e^{\omega_3\rho(x-s)}\int_x^s r(t)\delta_3[\rho(s-t)]e^{\omega_3\rho(t-x)}v(x, t, \rho)dt. \end{aligned}$$

Since $\tau(x, \rho)$ and its first two derivatives are continuous in a closed interval, we have $|\tau(x, \rho)|, |\tau'(x, \rho)|, |\tau''(x, \rho)| < K/|\rho|$, where K is independent of x and of ρ . Also $|\tau(x, \rho) - 3f(x)/\rho| < K/|\rho|^2$, where K has been increased, if necessary. Hence, letting $M(x, \rho)$ be the maximum of $|v(x, s, \rho)|$ for $x \leq s \leq \pi$, we have $M(x, \rho) < K' + K''M(x, \rho)/|\rho|^2$, where K' and K'' are both independent of x and of ρ . Here, of course, we have restricted ρ to the sector S_1 . It follows that $v(x, s, \rho)$ is an E -function if $|\rho|$ is sufficiently large and ρ is in S_1 .

We need also the asymptotic forms of $\sigma'_s(x, s)$ and $\sigma''_s(x, s)$. These are found from the equations obtained from the equation of Theorem 14 by differentiation with respect to s . We incorporate them in the statement of

THEOREM XV. *If equations (8) are satisfied, then*

$$\begin{aligned} \sigma(x, s) = & -v(\rho)u(s)/\rho^2 + \delta_1[\rho(s-x)]f(x)/\rho + e^{\omega_3\rho(s-x)}\rho^{-2}E(x, s, \rho), \\ \sigma'_s(x, s) = & -v(\rho)u'(s)/\rho^2 - \delta_3[\rho(s-x)]f(x) + e^{\omega_3\rho(s-x)}\rho^{-1}E(x, s, \rho), \\ \sigma''_s(x, s) = & -v(\rho)u''(s)/\rho^2 + \rho\delta_2[\rho(s-x)]f(x) + e^{\omega_3\rho(s-x)}E(x, s, \rho), \end{aligned}$$

provided $x \leq s \leq \pi$, $|\rho|$ is large, and ρ is in S_1 .

We need also the asymptotic form of the t -integral in equation (11). This is given in

THEOREM XVI.

$$\begin{aligned} \int_x^\pi f(t)W_{2\pi}(u_3)dt = & 3\beta_{20}f(\pi)/\rho + 3\beta_{21}f'(\pi)/\rho + e^{\omega_3\rho(\pi-x)}\rho^{j-2}E(x, \rho) \\ & - f(x)[\beta_{22}\rho^2\delta_2[\rho(\pi-x)] - \beta_{21}\rho\delta_3[\rho(\pi-x)] + \beta_{20}\delta_1[\rho(\pi-x)]]/\rho. \end{aligned}$$

This form is obtained by using the special case of Theorem 2 in which

$\alpha_{12} = 0$ and $\alpha_{11}\alpha_{30} - \alpha_{10}\alpha_{31} = 1$. Taking the asymptotic forms there given, we have

$$\int_x^\pi f(t) W_{2\pi}(u_s) dt = \int_x^\pi f(t) [\beta_{22}\rho^2\delta_1[\rho(\pi-t)] - \beta_{21}\rho\delta_2[\rho(\pi-t)] \\ + \beta_{20}\delta_3[\rho(\pi-t)]] dt + \rho^{j-2} \int_x^\pi f(t) e^{\omega_3\rho(\pi-t)} E(t, \rho) dt.$$

Integrating by parts twice the first integral on the right-hand side of this equation gives an equation equivalent to the one in the statement of the theorem.

We are now ready to insert the results of Theorems 13, 15, and 16 into equation (11). Using at the same time the conjugate property of the integrand in equation (11), we obtain, after making simple reductions,

$$I_n(x) = \frac{1}{\pi i_n} \int_{\gamma_n} \left[\tau(x, \rho) - \frac{\nu(\rho)u(x)}{\rho^2\Delta(\rho)} \{\Delta(\rho) + 9\rho W_{2\pi}(u)\} \right. \\ \left. + \frac{2\gamma u(x)}{\Delta(\rho)} \{\beta_{20}f(\pi) + \beta_{21}(f'(\pi))\} + \frac{u(x)}{\Delta(\rho)} e^{\omega_3\rho(\pi-x)} \rho^{j-1} E(x, \rho) \right] d\rho,$$

where γ_n is the part of γ_n in S_1 .

But $\tau(x, \rho) = 3f(x)/\rho + E(x, \rho)/\rho^4$, $u(x) = \rho^m e^{\omega_3\rho x} E(x, \rho)$, $\Delta(\rho) + 9\rho W_{2\pi}(u) = 2\gamma D_a \rho^3$ and $\nu(\rho)/\rho^2 = E(\rho)/\rho^3$. Remembering also inequality (5), we see that $I_n(x) = f(x) + \epsilon_n(x)$, where $\epsilon_n(x)$ tends uniformly to zero as n becomes infinite. We sum this up in

THEOREM XVII. *If*

- 1) $f(x)$ and $r(x)$ possess derivatives of all orders in an interval of which $x = 0$ is an interior point,
- 2) $r(x)$ and $f'(x)$ are continuous for $0 \leq x \leq \pi$,
- 3) $f(x)$ satisfies equations (8), and
- 4) the series defining $\tau(x, \rho)$ converges uniformly in the interval of hypothesis 1), then the formal series for $f(x)$ converges uniformly to $f(x)$ on every closed interval $0 \leq x \leq \beta < \pi$ interior to the interval mentioned in hypothesis 1).

UNIVERSITY OF IOWA.

ON THE DIFFERENTIATION OF INFINITE CONVOLUTIONS.

By AUREL WINTNER.

The object of the present note is an elementary theorem on term-by-term differentiation which, when applied to infinite convolutions of distribution functions,† implies results of the following type:

*If at least one term $\sigma_k = \sigma_k(x)$, $-\infty < x < +\infty$, of the convergent infinite convolution $\sigma_1 * \sigma_2 * \dots$ has an absolutely integrable and bounded second derivative, then, as $n \rightarrow \infty$, the continuous density of $\sigma_1 * \sigma_2 * \dots * \sigma_n$ tends to that of $\sigma_1 * \sigma_2 * \dots$ for every x .*

The assumption that a σ_k has an absolutely integrable and bounded second derivative does not *presuppose* that the Fourier transforms of the densities of the finite and infinite convolutions vanish at infinity more strongly than $o(|t|^{-1})$; and $o(|t|^{-1})$ is an estimate which does not suffice for the absolute integrability of these Fourier transforms.

It will be convenient to consider open intervals only. The classical theorem of Dini on term-by-term differentiation states that if a sequence $\{f_n(x)\}$ of differentiable functions is convergent and the sequence $\{f'_n(x)\}$ is uniformly convergent in an interval (a, b) , then $\lim f_n(x)$ is differentiable and its derivative is equal to $\lim f'_n(x)$ at every point of (a, b) . This theorem and its usual analogues introduce an assumption regarding the *convergence of the sequence of the derivatives*. For the case of infinite convolutions, a criterion is necessary which is free of such an assumption. A criterion of this type is suggested by, and effectively may be deduced from, the theory of convex functions. It will, however, be convenient to present the proof in a somewhat modified form. One advantage of this presentation is that the proof may easily be extended to the case of more than one variable. The criterion is independent of the Lebesgue theory.

A sequence of functions will be said to be of uniformly bounded variation in (a, b) if the total variation of the n -th function in (a, b) is less than a number which is independent of n . Under this condition the sequence is uniformly bounded in the interval if it is bounded at one point of the interval. The criterion in question runs now as follows:

If a convergent sequence $\{f_n(x)\}$ of differentiable functions is such that

† As to terminology, cf. a joint paper of B. Jessen and the present author, appearing in the *Transactions of the American Mathematical Society*.

$\{f'_n(x)\}$ is uniformly bounded and of uniformly bounded variation in (a, b) , then

- (i) $\{f_n(x)\}$ is uniformly convergent in (a, b) ;
- (ii) $f(x) = \lim f_n(x)$ has at every point of (a, b) a right-hand and a left-hand derivative, and both derivatives are bounded in (a, b) ;
- (iii) $f'_n(x) \rightarrow f'(x)$ at every x for which $f'(x)$ exists;
- (iiii) $f'(x)$ exists with the possible exception of a set of points x which is at most enumerable.

It may be mentioned that $f'_n(x)$ is continuous; in fact, a function of bounded variation cannot have a discontinuity of the second kind and a derivative cannot have a discontinuity of the first kind.

Since $\{f'_n(x)\}$ is uniformly bounded, $\{f_n(x)\}$ satisfies a uniform Lipschitz condition

$$|f_n(x_1) - f_n(x_2)| < M |x_1 - x_2|,$$

where M is independent of x_1, x_2 , and n . Now a sequence of functions which satisfy a uniform Lipschitz condition is, according to a theorem of Arzelà, uniformly convergent in (a, b) if it is convergent on a dense set of (a, b) . This proves (i). It is seen that it was not necessary to suppose the convergence of $\{f_n(x)\}$ at every point of (a, b) .

Every uniformly bounded sequence of monotone non-decreasing functions contains an everywhere convergent subsequence; this is a well-known theorem of Helly. It is obvious that if a sequence of functions is uniformly bounded and of uniformly bounded variation, then it may be represented as the difference of two sequences each of which consists of monotone non-decreasing functions which are uniformly bounded. Hence, if a sequence of functions is uniformly bounded and of uniformly bounded variation in (a, b) , then it contains a subsequence which is convergent at every point of (a, b) .

Let $\{f'_{m_n}(x)\}$ be a convergent subsequence of $\{f'_n(x)\}$. Put $g_n(x) = f_{m_n}(x)$ and let $g'_n(x) \rightarrow G(x)$, so that

$$\int_c^x g'_n(t) dt \rightarrow \int_c^x G(t) dt,$$

since $\{g'_n(x)\}$ is uniformly bounded. On the other hand,

$$\int_c^x g'_n(t) dt = g_n(x) - g_n(c) \rightarrow f(x) - f(c),$$

since $f_n(x) \rightarrow f(x)$. Consequently,

$$f(x) - f(c) = \int_c^x G(t) dt.$$

This implies (ii) and (iii), since $G(x)$ is the limit of functions of uniformly bounded variation and is therefore of bounded variation. It is seen that if $f'(x)$ exists at $x = x_0$, then $f'(x_0) = G(x_0)$, so that $g'_n(x_0) \rightarrow f'(x_0)$ holds whenever $\{g'_n(x)\}$ is a subsequence of $\{f'_n(x)\}$ which is convergent at every point of (a, b) .

Suppose finally that (iii) is false, i. e., that there is a point x_0 such that $f'(x_0)$ exists but $f'_n(x_0) \rightarrow f'(x_0)$ does not hold. Since $\{f'_n(x_0)\}$ is a bounded sequence of numbers, it contains a subsequence $\{h'_n(x_0)\}$ such that $h'_n(x_0) \rightarrow l$, where $l \neq f'(x_0)$. Consider now the corresponding sequence of functions $\{h'_n(x)\}$; it is a subsequence of $\{f'_n(x)\}$, hence uniformly bounded and of uniformly bounded variation. Thus the sequence $\{h'_n(x)\}$ contains a subsequence $\{g'_n(x)\}$ which is convergent at every point of (a, b) . This subsequence of $\{h'_n(x)\}$ is a subsequence of $\{f'_n(x)\}$ and tends therefore at $x = x_0$ to $f'(x_0)$ in virtue of the last remark of the previous paragraph. On the other hand, every subsequence of $\{h'_n(x_0)\}$ tends to l in virtue of $h'_n(x_0) \rightarrow l$. Consequently $f'(x_0) = l$. This completes the proof, since $f'(x_0) \neq l$ by hypothesis.

The enumerable set mentioned in (iiii) may actually exist and it may even be dense in (a, b) . In fact, it is easy to see that every convex function $f(x)$ satisfying a Lipschitz condition may be approximated by a sequence $\{f_n(x)\}$ of differentiable functions for which $\{f'_n(x)\}$ is uniformly bounded and of uniformly bounded variation. On the other hand, there exist convex functions which satisfy a Lipschitz condition but have a dense set of corners.

The theorem yields a result, viz. (iii), also in cases where the *existence* of the derivative of the limit function is presupposed or obvious for every x . An instance of this situation is the case of infinite convolutions.

Let $\rho(x)$ be a distribution function possessing an absolutely integrable and bounded second derivative. Then $\rho'(x) \leq C$, where

$$C = \int_{-\infty}^{+\infty} |\rho''(x)| dx.$$

Since $\rho'(x)$ and $\rho''(x)$ are bounded Baire functions, they are integrable in the Stieltjes-Lebesgue sense with respect to any distribution function $\tau(x)$. The convolution $\rho * \tau$ has the continuous density

$$\int_{-\infty}^{+\infty} \rho'(x-t) d\tau(t) \quad (0 \leq \rho' \leq C)$$

which is not greater than C , a bound which is independent of x and of the distribution function τ . Furthermore, the total variation of the density of $\rho * \tau$ is

$$\int_{-\infty}^{+\infty} |d\{\int_{-\infty}^{+\infty} \rho'(x-t) d\tau(t)\}/dx| dx = \int_{-\infty}^{+\infty} |\int_{-\infty}^{+\infty} \rho''(x-t) d\tau(t)| dx,$$

which is not greater than

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\rho''(x-t)| dx d\tau(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\rho''(s)| ds d\tau(t) = C \cdot 1,$$

where C is independent of the distribution function τ . Hence if σ_1 is a distribution function possessing an absolutely integrable and bounded second derivative and if $\sigma_2, \sigma_3, \dots$ are arbitrary distribution functions, it follows, by placing $\rho = \sigma_1$ and $\tau = \sigma_2 * \dots * \sigma_n$, that the sequence $\{f'_n(x)\}$, where $f_n = \sigma_1 * \dots * \sigma_n$, exists and is uniformly bounded and of uniformly bounded variation. Finally, if the infinite convolution $\sigma_1 * \sigma_2 * \dots$ is convergent, then it possesses a continuous density, since $\rho * \tau = \sigma_1 * \tau$ has a continuous density for any τ , so that one may choose $\tau = \sigma_2 * \sigma_3 * \dots$.

It is clear from the proof that the assumption regarding σ_1 may be replaced by a somewhat weaker one, and that higher derivatives of the infinite convolution $\sigma_1 * \sigma_2 * \dots$ may be similarly treated.

A convergence theory of infinite convolutions has been developed in the joint paper of Jessen and the present author, referred to above. There is an *explicit* sufficient convergence criterion which is of interest insofar as it applies also in cases where the distribution functions occurring in the infinite convolution do not possess finite *second* moments:

If $\sum_{n=1}^{\infty} M_n < +\infty$, where

$$M_n = \int_{-\infty}^{+\infty} |x| d\sigma_n(x),$$

then the infinite convolution $\sigma_1 * \sigma_2 * \dots$ is absolutely convergent.

In fact, $M_n < +\infty$ implies that the Fourier-Stieltjes transform

$$L(t; \sigma_n) = \int_{-\infty}^{+\infty} e^{itx} d\sigma_n(x)$$

of σ_n has for every t a continuous first derivative of absolute value $\leq M_n$. Hence $|L(t; \sigma_n) - 1| \leq |t| M_n$ in virtue of $L(0; \sigma_n) = 1$. It follows therefore from the convergence of the series $M_1 + M_2 + \dots$ that the infinite product $L(t; \sigma_1)L(t; \sigma_2) \dots$ is absolutely and uniformly convergent in every finite t -interval. This means that the infinite convolution $\sigma_1 * \sigma_2 * \dots$ is absolutely convergent.

POLYNOMIALS OF BEST APPROXIMATION ASSOCIATED WITH CERTAIN PROBLEMS IN TWO DIMENSIONS.

By W. H. McEWEN.

1. *Introduction.* Let $u(x, y)$ be a function which is defined and continuous and possesses continuous partial derivatives of the 1st and 2nd orders throughout a square region of the xy -plane $a \leq x, y \leq b$. Let C be a closed curve lying wholly within the square, and let J denote the region bounded by C . Then, if it is a question of approximating to $u(x, y)$ throughout the region J by means of polynomials of the form

$$P_{mn}(x, y) = \sum_{i,j}^{m,n} a_{ij} x^i y^j,$$

the problem becomes definite only when a measure of best approximation is determined upon. In this paper we shall consider in turn two different situations as regards the function u and the curve C , designated below as problems A and B respectively, and in each shall define a measure of best approximation and obtain theorems on the convergence of P_{mn} as m, n both become infinite.

Problem A. This problem is characterised by two additional assumptions that we make respecting C and u :

(1) C is an algebraic curve, and hence may be represented by the equation

$$c(x, y) = 0,$$

where $c(x, y)$ is a polynomial of some specified degrees m', n' .

(2) $u(x, y)$ vanishes identically on C , i. e. $u(\alpha, \beta) \equiv 0$, where (α, β) represents a variable point on C .

For the determination of $P_{mn}(x, y)$, the polynomial of best approximation to u of degrees m, n , we shall use

Criterion A. $P_{mn}(x, y)$ must vanish identically on C , $P_{mn}(\alpha, \beta) \equiv 0$, and must give at the same time a minimum value to the expression

$$\iint_J |\nabla^2(u - P_{mn})|^r dx dy, \quad \nabla^2 w \equiv \partial^2 w / \partial x^2 + \partial^2 w / \partial y^2,$$

in comparison with all other polynomials of like degrees which vanish identically on C , r being any given constant > 0 .

Our special concern will be of course to prove that under suitable addi-

tional hypotheses the polynomials P_{mn} will converge uniformly throughout J to the value of u as m and n both become infinite. Denoting the value of $\nabla^2 u$ by $R(x, y)$, it is clear from the manner in which P_{mn} is defined that the problem could be regarded also from another standpoint, namely that of furnishing an approximation to the solution of a given differential system $\nabla^2 u = R(x, y)$, $u = 0$ on C . However, if this point of view were adopted it would be necessary to introduce into the discussion certain questions of an incidental nature, relating to the extension of the definition of the solution u to apply in that region of the square which lies outside of J . By assuming in the first place that u is defined throughout the square, we have been able to avoid these additional questions and thereby to focus attention more fully upon the processes involved in the proofs of convergence proper.

Problem B. In this case we shall discard the assumptions made in A, so that for the present at least, we may consider C as any closed curve lying wholly within the square, and u as the function described in the first paragraph and taking on arbitrary values on C . For a given pair of positive integers m, n the polynomial of best approximation P_{mn} of degrees m, n will be defined by

Criterion B. P_{mn} must give a minimum value to the expression

$$\iint_J |\nabla^2(u - P_{mn})|^r dx dy + \lambda \max_{\text{on } C} |u(\alpha, \beta) - P_{mn}(\alpha, \beta)|^s,$$

in comparison with all other polynomials of like degrees, r, s and λ being any given constants > 0 , and (α, β) the coördinates of a variable point on C .

For each problem we shall give two proofs of convergence, one based on Hölder's inequality and applicable only when $r > 1$, and the other depending on Markoff's theorem on the derivative of a polynomial in two dimensions and applicable generally when r is any real number > 0 . By way of comparison it will be seen that for cases in which $r > 1$ the first method requires a less restrictive hypothesis than the second. The writer has considered already situations in one dimension corresponding to problems A and B,* while Kryloff † has treated a problem similar to A but for approximating sums connected with the method of Ritz.

* W. H. McEwen, "Problems of closest approximation connected with the solution of linear differential equations," *Transactions of the American Mathematical Society*, vol. 33 (1931), pp. 979-997; "On the approximate solution of linear differential equations with boundary conditions," *Bulletin of the American Mathematical Society*, vol. 38 (1932), pp. 887-894.

† N. Kryloff, "Application de la méthode de l'algorithme variationnel à la solution

In connection with these problems it is of interest to note that the results obtained in this paper, and also the reasoning used with only slight modifications, are valid when the measures of best approximation are altered to the extent that the double integral of the r -th power of $|\nabla^2(u - P_{mn})|$ is replaced by the term

$$\lambda' \max \text{ in } J |\nabla^2(u - P_{mn})|^r,$$

r and λ' being any given positive constants. This statement can be made even stronger by asserting that for the case $0 < r \leq 1$ the hypotheses demanded by our theorems II and IV for convergence can be lightened to agree exactly with that required when $r > 1$.

2. *Preliminary discussion.* In anticipation of later needs we shall develop next some results concerning the simultaneous approximation of an arbitrary function $v(x, y)$ and its partial derivatives of first and second order, by means of polynomials and their corresponding derivatives.

Let $v(x, y)$ be defined throughout the square $a \leq x, y \leq b$. For simplicity in exposition we shall take this square to be $-1 \leq x, y \leq 1$, although the results obtained apply equally to the more general case. Suppose further that $v(x, y)$ and its partial derivatives of 1st and 2nd order are continuous throughout the square.

By means of the transformation $x = \cos \theta, y = \cos \phi$, we can put v in the form of a periodic function

$$v(\cos \theta, \cos \phi) = \bar{v}(\theta, \phi)$$

having the period 2π in both its arguments θ and ϕ , and thus having the entire $\theta\phi$ -plane as its region of definition. Then, by expressing \bar{v} and its derivatives with respect to θ and ϕ in terms of v and its derivatives with respect to x and y , it is readily seen that the hypothesis made in the paragraph above concerning $v(x, y)$ will carry over automatically to $\bar{v}(\theta, \phi)$. Hence for all values of θ and ϕ , \bar{v} and its partial derivatives of 1st and 2nd order are continuous.

But for a periodic function which is continuous, such as $\bar{v}(\theta, \phi)$, Mickelson* has shown that for every pair of positive integers m and n there exists a trigonometric sum of orders m, n

approchée des équations différentielles aux dérivées partielles du type elliptique," *Bulletin de l'Académie de l'U. R. S. S.*, 1930.

* E. L. Mickelson, "On the approximate representation of a function of two variables," *Transactions of the American Mathematical Society*, vol. 33 (1931), pp. 759-781; p. 76, Theorem II. In this connection see also C. E. Wilder, "On the degree of approximation to discontinuous functions by trigonometric sums," *Rendiconti del*

$$T_{mn}(\theta, \phi) = \sum_{i,j}^{m,n} [A_{ij} \cos i\theta \cos j\phi + B_{ij} \cos i\theta \sin j\phi \\ + C_{ij} \sin i\theta \cos j\phi + D_{ij} \sin i\theta \sin j\phi]$$

such that

$$|\bar{v}(\theta, \phi) - T_{mn}(\theta, \phi)| \leq K_1 \omega(1/m + 1/n)$$

for all values of θ and ϕ , K_1 being a constant independent of m and n , and $\omega(\delta)$ being the modulus of continuity of \bar{v} . It will be well to observe at this point that for functions which are uniformly continuous, such as those with which we will be concerned, $\lim_{\delta \rightarrow 0} \omega(\delta) = 0$. The function T_{mn} may be obtained, for example, by making an extension to two dimensions of Jackson's approximating function.* The result is

$$T_{mn}(\theta, \phi) = I_{pq}(\theta, \phi) = h_{pq} \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} \bar{v}(\theta + 2\lambda, \phi + 2\mu) F_{pq}(\lambda, \mu) d\lambda d\mu,$$

where

$$F_{pq}(\theta, \phi) = \left[\frac{(\sin p\lambda)(\sin q\mu)}{(p \sin \lambda)(q \sin \mu)} \right]^4,$$

$$1/h_{pq} = \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} F_{pq}(\lambda, \mu) d\lambda d\mu,$$

and p and q are two integers such that $2p - 2 \leq m \leq 2p$ and $2q - 2 \leq n \leq 2q$.

Letting $\xi = \theta + 2\lambda$ and $\eta = \phi + 2\mu$ and substituting under the integral signs for λ, μ , and making use of the fact that the integrand has the period 2π in both the variables ξ and η , we get

$$T_{mn}(\theta, \phi) = I_{pq}(\theta, \phi) = h_{pq} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \bar{v}(\xi, \eta) \Phi(\xi - \theta) \Phi(\eta - \phi) d\xi d\eta$$

where

$$\Phi(w) = \frac{1}{2} [(\sin pw/2)/(p \sin w/2)]^4.$$

On differentiating this result with respect to θ , and replacing $\partial \Phi(\xi - \theta)/\partial \theta$ by its equal $-[\partial \Phi(\xi - \theta)/\partial \xi]$, and then integrating the resulting expression by parts, we get

$$\frac{\partial}{\partial \theta} T_{mn}(\theta, \phi) = h_{pq} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \xi} \bar{v}(\xi, \eta) \cdot \Phi(\xi - \theta) \Phi(\eta - \phi) d\xi d\eta,$$

which is precisely the I_{pq} -function associated with $\partial \bar{v}/\partial \theta$. Then, since $\partial \bar{v}/\partial \theta$ is

Circolo Matematico di Palermo, vol. 39 (1915), pp. 345-361; p. 358, Theorem X, in which it is shown that if the given function satisfies a Lipschitz condition the absolute value of the error will not exceed a constant multiple of $(1/m + 1/n)$.

* D. Jackson, "The theory of approximations," *American Mathematical Society Colloquium Publications*, New York, 1930, p. 3.

itself continuous, there must exist, according to Mickelson's result, a constant K_2 independent of m and n such that

$$|\partial \bar{v}/\partial \theta - \partial T_{mn}/\partial \theta| \leq K_2 \omega(1/m + 1/n)$$

for all values of θ and ϕ . Similarly we can show that the remaining derivatives of $\bar{v} - T_{mn}$ of 1st and 2nd order have upper bounds which are constant multiples of $\omega(1/m + 1/n)$.

Furthermore, since $\bar{v}(\theta, \phi) = v(\cos \theta, \cos \phi)$ is an even function of both θ and ϕ , the function T_{mn} will necessarily be even. Hence, on changing back again to the variables x, y , we are led at once to a polynomial $p_{mn}(x, y)$ of degrees m, n , while the region of approximation becomes again the square $-1 \leq x, y \leq 1$. Moreover from the identities

$$\begin{aligned} \bar{v}(\theta, \phi) - T_{mn}(\theta, \phi) &= v(x, y) - p_{mn}(x, y), \\ \frac{\partial}{\partial \theta}[\bar{v}(\theta, \phi) - T_{mn}(\theta, \phi)] &= (1 - x^2)^{1/2} \frac{\partial}{\partial x}[v(x, y) - p_{mn}(x, y)], \\ \frac{\partial^2}{\partial \theta^2}[\bar{v}(\theta, \phi) - T_{mn}(\theta, \phi)] &= (1 - x^2) \frac{\partial^2}{\partial x^2}[v(x, y) - p_{mn}(x, y)] \\ &\quad + x \frac{\partial}{\partial x}[v(x, y) - p_{mn}(x, y)], \text{ etc.} \end{aligned}$$

it is clear that if we restrict our attention to a region J which lies wholly within the square, so that $(1 - x^2)^{1/2}$ has a positive lower bound, then the quantities $(\partial^{i+j}/\partial x^i \partial y^j)[v(x, y) - p_{mn}(x, y)]$, $(i + j = 0, 1, 2)$, will have upper bounds in J which are constant multiples of $\omega(1/m + 1/n)$.

Furthermore, if the hypothesis regarding v is extended so that v and its partial derivatives of orders $1, 2, \dots, k$ ($k > 2$) are continuous throughout the square, then by an appropriate generalization of the function I_{pq} ,* the argument given above can be used to prove that the expressions

$$\partial^{i+j}(v - p_{mn})/\partial x^i \partial y^j, \quad (i + j = 0, 1, 2)$$

have upper bounds in J which are constant multiples of

$$(1/m + 1/n)^k \Omega(1/m + 1/n),$$

where $\Omega(\delta)$ is the greatest of the moduli of continuity associated with the k -th order derivatives of v .

The results obtained in this section thus far may be summarized in the following two theorems.

* See Mickelson, *loc. cit.*, pp. 766-768.

THEOREM A. If $v(x, y)$ and its partial derivatives of the 1st and 2nd orders are continuous throughout the square $a \leq x, y \leq b$, then for every pair of positive integers m and n there exists a polynomial $p_{mn}(x, y)$ of degrees m, n , and a positive constant K independent of m and n , such that the relations

$$|\partial^{i+j}(v - p_{mn})/\partial x^i \partial y^j| \leq K\omega(1/m + 1/n), \quad (i + j = 0, 1, 2)$$

hold uniformly throughout any closed region J which lies wholly within the square.

THEOREM B. If $v(x, y)$ and its partial derivatives of orders $1, 2, \dots, k$ are continuous throughout $a \leq x, y \leq b$, then, for every pair of positive integers m, n , there exists a polynomial $p_{mn}(x, y)$ of degrees m, n , and a positive constant K' independent of m and n , such that the relations

$$|\partial^{i+j}(v - p_{mn})/\partial x^i \partial y^j| \leq K'(1/m + 1/n)^k \cdot \Omega(1/m + 1/n) \quad (i + j = 0, 1, 2)$$

hold uniformly throughout the region J .

As yet we have not considered the questions of existence and uniqueness in relation to our polynomials of best approximation. It is not difficult to show that in both problems polynomials P_{mn} as defined by the respective criteria do exist, and moreover when $r > 1$ are uniquely determined. Thus in problem A where C is an algebraic curve represented by $c(x, y) = 0$, and P_{mn} is required to vanish identically on C , we can write

$$P_{mn}(x, y) = \sum_{i,j}^{m-m', n-n'} b_{ij} \psi_{ij}(x, y),$$

where $\psi_{ij}(x, y) = c(x, y)x^i y^j$. Then since no polynomial which so vanishes can be harmonic in J (unless it be identically zero there), it follows that

$$\nabla^2 P_{mn} = \sum_{i,j}^{m-m', n-n'} b_{ij} \nabla^2 \psi_{ij} \text{ cannot vanish identically in } J \text{ and hence that it may}$$

be regarded as a linear combination of functions $\nabla^2 \psi_{ij}$ which are linearly independent in J . On the basis of this result the existence and uniqueness theorems can be proved by the use of an argument exactly similar to that used in the one dimensional problem.* By a suitable modification of the wording the same type of argument would suffice also in problem B.

3. *Problem A. Convergence in the special case $r > 1$.* Consider the function $v(x, y) = u(x, y)/c(x, y)$. Let $p_{m-m', n-n'}$ be a polynomial of degrees $m-m', n-n'$, arbitrary for the moment, and let $\epsilon > 0$ be such that the relations

* See the writer's first paper, *loc. cit.*

$$(1) \quad \left| \partial^{i+j}(v - p_{m-m', n-n'}) / \partial x^i \partial y^j \right| \leq \epsilon \quad (i+j=0, 1, 2)$$

hold uniformly throughout J . Ultimately we shall assume that v satisfies the hypothesis of Theorem A, so that ϵ may be taken to be

$$K\omega[1/(m-m') + 1/(n-n')]$$

and hence $\lim_{m, n \rightarrow \infty} \epsilon = 0$.

Let $\pi_{mn} = cp_{m-m', n-n'}$. Then π_{mn} is a polynomial of degrees m, n , and furthermore

$$u - \pi_{mn} = c(v - p_{m-m', n-n'}),$$

$$\frac{\partial}{\partial x}(u - \pi_{mn}) = (v - p_{m-m', n-n'}) \frac{\partial c}{\partial x} + c \frac{\partial}{\partial x}(v - p_{m-m', n-n'}), \text{ etc.}$$

From these relations and (1) it is clear that the upper bounds in J of $\left| \partial^{i+j}(u - \pi_{mn}) / \partial x^i \partial y^j \right|$, $(i+j=0, 1, 2)$ are expressible linearly in terms of ϵ and the upper bounds of c and its derivatives. Hence there must exist a constant B independent of m and n to satisfy the inequalities

$$(2) \quad \left| \partial^{i+j}(u - \pi_{mn}) / \partial x^i \partial y^j \right| \leq B\epsilon \quad (i+j=0, 1, 2)$$

uniformly throughout J . In particular then

$$(3) \quad |\nabla^2(u - \pi_{mn})| \leq 2B\epsilon.$$

Now the polynomial of best approximation P_{mn} of degrees m, n is defined so as to vanish on C and at the same time to minimize the expression

$$\iint_J |\nabla^2(u - P_{mn})|^r dx dy$$

in comparison with all other polynomials of like degrees which so vanish. Such another polynomial is π_{mn} . Hence, by virtue of this and (3), we can write

$$(4) \quad \iint_J |\nabla^2(u - P_{mn})|^r dx dy$$

$$\leq \iint_J |\nabla^2(u - \pi_{mn})|^r dx dy \leq A(2B\epsilon)^r,$$

A being the area of the region J .

Let $G(x, y; \xi, \eta)$ be the Green's function of two dimensions associated with the homogeneous differential system $\nabla^2 w = 0$, $w = 0$ on C . Then, since u and P_{mn} both vanish identically on C , it is possible to write

$$u(x, y) = \iint_J G(x, y; \xi, \eta) \nabla^2 u(\xi, \eta) d\xi d\eta,$$

$$P_{mn}(x, y) = \iint_J G(x, y; \xi, \eta) \nabla^2 P_{mn}(\xi, \eta) d\xi d\eta,$$

and therefore also

$$u - P_{mn} = \iint_J G(x, y; \xi, \eta) \nabla^2 [u(\xi, \eta) - P_{mn}(\xi, \eta)] d\xi d\eta.$$

The function G is not bounded in J (becoming infinite as

$$\log \sqrt{(x - \xi)^2 + (y - \eta)^2}$$

at the point (ξ, η)), but nevertheless the double integrals over J of $|G|$ and $|G|^{r/(r-1)}$ are finite in value. Hence, the number r being > 1 , it is possible to apply Hölder's inequality to this last relation and so obtain the result

$$|u - P_{mn}| \leq \left[\iint_J |G|^{r/(r-1)} d\xi d\eta \right]^{1-1/r} \cdot \left[\iint_J |\nabla^2 (u - P_{mn})|^r d\xi d\eta \right]^{1/r}.$$

The first factor occurring on the right is merely a constant, whereas the second is bounded as shown in (4). Hence there must exist a constant D independent of m and n to satisfy the relation

$$|u - P_{mn}| \leq D\epsilon$$

uniformly throughout J .

If we assume now that $v = u/c$ satisfies the hypothesis of Theorem A so that ϵ may be chosen to make $\lim_{m,n \rightarrow \infty} \epsilon = 0$, then it is certain from this last result that P_{mn} will converge uniformly in J to the value of u as m and n both become infinite.

Likewise we can show that the partial derivatives of the 1st and 2nd orders converge. For we can write

$$\frac{\partial^{i+j}}{\partial x^i \partial y^j} (u - P_{mn}) = \iint_J \frac{\partial^{i+j} G}{\partial x^i \partial y^j} \cdot \nabla^2 (u - P_{mn}) d\xi d\eta, \quad (i + j = 1, 2),$$

and so by Hölder's inequality and (4) obtain

$$|\partial^{i+j} (u - P_{mn}) / \partial x^i \partial y^j| \leq D'\epsilon, \quad (i + j = 1, 2),$$

where D' is a constant independent of m and n . Thus we can state

THEOREM I. *In problem A in the case when $r > 1$, if the function*

$u(x, y)/c(x, y)$ satisfies the hypothesis of Theorem A, there will exist a positive quantity $\epsilon_{mn} = [1/(m - m') + 1/(n - n')]$, and a positive constant D_1 independent of m and n , such that the relations

$$|\partial^{i+j}(u - P_{mn})/\partial x^i \partial y^j| \leq D_1 \epsilon_{mn}, \quad (i + j = 0, 1, 2)$$

hold uniformly throughout J , and $\lim_{m, n \rightarrow \infty} \epsilon_{mn} = 0$.

4. *Problem A. Convergence in the general case when $r > 0$.* Let $F(x, y) = u(x, y) - \pi_{mn}(x, y)$, where π_{mn} is the polynomial described in the last section satisfying relations (2) and (3),

$$(2) \quad |\partial^{i+j}(F)/\partial x^i \partial y^j| \leq B\epsilon, \quad (i + j = 0, 1, 2),$$

$$(3) \quad |\nabla^2(F)| \leq 2B\epsilon.$$

Then the function F , like u , vanishes identically on C and hence there will exist for it a polynomial of best approximation Q_{mn} of degrees m, n (Criterion A). Moreover Q_{mn} will vanish identically on C and the double integral

$$\gamma = \iint_J |\nabla^2(F - Q_{mn})|^r dx dy$$

will be a minimum for polynomials of like degrees which so vanish. But 0 may be regarded as another such polynomial vanishing on C , and hence

$$\gamma \leq \iint_J |\nabla^2(F)|^r dx dy, \text{ and therefore by reason of (3),}$$

$$(6) \quad \gamma \leq A(2B\epsilon)^r.$$

Let δ be the maximum value of $|\nabla^2 Q_{mn}|$ in the region J , and let (x_0, y_0) be a point of J at which $|\nabla^2 Q_{mn}(x_0, y_0)| = \delta$. Then, since $\nabla^2 Q_{mn}$ is a polynomial of degrees not exceeding m, n , it follows as a consequence of Markoff's theorem that $|\partial \nabla^2 Q_{mn}/\partial x| \leq H\bar{m}^2\delta$ and $|\partial \nabla^2 Q_{mn}/\partial y| \leq H\bar{m}^2\delta$ throughout the region J , \bar{m} being the greater of the two numbers m and n , and H a constant depending on the region J^* and independent of m and n . In the light of these results and the mean value theorem we can write

$$|\nabla^2 Q_{mn}(x, y) - \nabla^2 Q_{mn}(x_0, y_0)| \leq [|x - x_0| + |y - y_0|] H\bar{m}^2\delta.$$

* In this connection it should be observed that certain broad requirements must be met by the region J in order to insure the applicability of Markoff's theorem. It will be sufficient to assume that J is a region for which there exists a positive constant h and a small angle $\theta \neq 0$ such that from every point of the boundary curve two line segments of lengths h and inclined at an angle θ with one another can be drawn belonging wholly to the region.

Now let us consider the square about the point (x_0, y_0) defined by the inequalities $|x - x_0| \leq 1/(4H\tilde{m}^2)$, $|y - y_0| \leq 1/(4H\tilde{m}^2)$. If j represent that part of the square which belongs to J , then throughout j , by virtue of the relation written above,

$$|\nabla^2 Q_{mn}(x, y) - \nabla^2 Q_{mn}(x_0, y_0)| \leq \delta/2,$$

and hence

$$(7) \quad |\nabla^2 Q_{mn}(x, y)| \geq \delta/2.$$

Let us assume for the moment that $\epsilon < \delta/(8B)$, so that $|\nabla^2(F)| \leq 2B\epsilon < \delta/4$. Then, by (7), $|\nabla^2(F - Q_{mn})| > \delta/4$ throughout j , and hence

$$\begin{aligned} \gamma &= \iint_j |\nabla^2(F - Q_{mn})|^r dx dy \\ &\geq \iint_j |\nabla^2(F - Q_{mn})|^r dx dy > [1/(4H\tilde{m}^2)^2] (\delta/4)^r. \end{aligned}$$

Therefore

$$\delta \leq 4[16H^2\tilde{m}^4\gamma]^{1/r}, \text{ and by (6)}$$

$$\delta \leq 4[16H^2\tilde{m}^4A]^{1/r}(2B\epsilon).$$

This result was proved on the basis of the assumption $\epsilon < \delta/(8B)$. However is this inequality does not hold, then $\delta \leq 8B\epsilon$. Hence in any case there will exist a constant E independent of m and n such that

$$(8) \quad \delta \leq E\tilde{m}^{4/r}\epsilon.$$

But the function Q_{mn} may be expressed in terms of the Green's function,

$$Q_{mn}(x, y) = \iint_j G(x, y; \xi, \eta) \nabla^2 Q_{mn}(\xi, \eta) d\xi d\eta.$$

Hence throughout J

$$|Q_{mn}| \leq \delta \iint_j |G(x, y; \xi, \eta)| d\xi d\eta = W\delta,$$

where W is a finite constant. Therefore, by (8),

$$|Q_{mn}| \leq WE\tilde{m}^{4/r}\epsilon.$$

From this and (2) it follows that

$$|F - Q_{mn}| \leq B\epsilon + WE\tilde{m}^{4/r}\epsilon \leq L\tilde{m}^{4/r}\epsilon,$$

where $L = (B + WE)$ is independent of m and n .

Now let us assume that $v = u/c$ satisfies the hypothesis of Theorem B with the integer k taken $\geq 4/r$, so that ϵ may be given the value $K'[1/(m - m') + 1/(n - n')]^k \Omega[1/(m - m') + 1/(n - n')]$. Then as m, n both become infinite $\bar{m}^{4/r} \epsilon$ will approach zero as a limit and therefore the quantity $|F - Q_{mn}|$ will converge to zero. But, as we have noted already, $F - Q_{mn}$ is identical with $u - P_{mn}$, where P_{mn} is the polynomial of best approximation to u . Thus we have proved that under the hypotheses stated P_{mn} converges to u . In a like manner it can be shown that the partial derivatives of the 1st and 2nd orders of P_{mn} converge to the respective derivatives of u . The results of this section are set forth in

THEOREM II. *In problem A in the general case $r > 0$, if the curve C is subject to the limitations imposed by the requirements of Markoff's theorem (see footnote *, p. 375), and if the function u/c satisfies the hypothesis of Theorem B with the integer k taken $\geq 4/r$, then there will exist a positive constant D_2 independent of m and n , and a positive quantity ϵ_{mn} such that the relations*

$$|\partial^{i+j}(u - P_{mn})/\partial x^i \partial y^j| \leq D_2 \epsilon_{mn}, \quad (i + j = 0, 1, 2)$$

hold uniformly throughout J , and furthermore, provided m and n maintain the same order of magnitude,

$$\lim_{m, n \rightarrow \infty} \epsilon_{mn} = 0.$$

An explicit formula for ϵ_{mn} is

$$\bar{m}^{4/r} [1/(m - m') + 1/(n - n')]^k \Omega[1/(m - m') + 1/(n - n')].$$

5. *Problem B. Convergence in the special case $r > 1$.* From this point on we discard the suppositions made in problem A that C be algebraic and that u vanish identically on C . Let p_{mn} be a polynomial of degrees m, n , which for the moment may be regarded as arbitrary, and let $\epsilon > 0$ satisfy the relations

$$(10) \quad |\partial^{i+j}(u - p_{mn})/\partial x^i \partial y^j| \leq \epsilon, \quad (i + j = 0, 1, 2)$$

uniformly throughout J . Then also

$$(11) \quad |\nabla^2(u - p_{mn})| \leq 2\epsilon.$$

* It must be understood here that m, n become infinite in such a way as to maintain at all times the same order of magnitude. That is, there must exist a constant α to satisfy the inequalities $1 \leq \bar{m}/m \leq \alpha$, $1 \leq \bar{m}/n \leq \alpha$. Then the coefficient of Ω in $\bar{m}^{4/r} \epsilon$ will not exceed $\alpha^{4/r} (m^{4/(kr)}/(m - m') + n^{4/(kr)}/(n - n'))^k$, a quantity which has a finite limit when $k \geq 4/r$.

But the polynomial of best approximation P_{mn} of degrees m, n (see Criterion B) is now defined to give a minimum value to the expression

$$\gamma = \int_J \int |\nabla^2(u - P_{mn})|^r dx dy + \lambda \max |u(\alpha, \beta) - P_{mn}(\alpha, \beta)|^s$$

in comparison with all other polynomials of degrees m, n , and therefore in particular with the polynomial p_{mn} . Hence

$$\gamma \leq \int_J \int |\nabla^2(u - p_{mn})|^r dx dy + \lambda \max |u(\alpha, \beta) - p_{mn}(\alpha, \beta)|^s,$$

and therefore, by virtue of (10) and (11),

$$\gamma \leq A(2\epsilon)^r + \lambda\epsilon^s.$$

Ultimately ϵ will be made to approach zero and so at this point we may assume that $2\epsilon < 1$. Then if q denote the smaller of the two numbers r, s

$$\gamma \leq (A + \lambda)(2\epsilon)^q.$$

But each term of γ is ≥ 0 and hence each $\leq \gamma$, and therefore

$$(12) \quad \int_J \int |\nabla^2(u - P_{mn})|^r dx dy \leq (A + \lambda)(2\epsilon)^q,$$

$$(13) \quad \max |u(\alpha, \beta) - P_{mn}(\alpha, \beta)| \leq [(A + \lambda)(2\epsilon)^q/\lambda]^{1/s}.$$

The function u may be expressed in terms of the Green's function,

$$u(x, y) = \int_J \int G(x, y; \xi, \eta) \nabla^2 u(\xi, \eta) d\xi d\eta + \phi(x, y),$$

where $\phi(x, y)$ is a function which is harmonic in the region J and which, on the boundary C , takes on the same values as does $u(x, y)$, i. e. $\phi(\alpha, \beta) = u(\alpha, \beta)$. So also,

$$P_{mn}(x, y) = \int_J \int G(x, y; \xi, \eta) \nabla^2 P_{mn}(\xi, \eta) d\xi d\eta + \psi(x, y),$$

where $\psi(x, y)$ is harmonic in J and $\psi(\alpha, \beta) = P_{mn}(\alpha, \beta)$. Then

$$u - P_{mn} = \int_J \int G \cdot \nabla^2(u - P_{mn}) d\xi d\eta + [\phi(x, y) - \psi(x, y)].$$

The number r being > 1 , Hölder's inequality can be applied to the first term on the right to give

$$\begin{aligned} & \left| \int_J \int G \cdot \nabla^2(u - P_{mn}) d\xi d\eta \right| \\ & \leq \left[\int_J \int |G|^{r/(r-1)} d\xi d\eta \right]^{1-1/r} \cdot \left[\int_J \int |\nabla^2(u - P_{mn})|^r d\xi d\eta \right]^{1/r}, \end{aligned}$$

from which it follows, by virtue of (12), that a constant M independent of m and n can be found such that

$$\left| \int_J \int G \cdot \nabla^2(u - P_{mn}) d\xi d\eta \right| \leq M\epsilon^{q/r}.$$

On the other hand the second term $(\phi - \psi)$, being harmonic in J , will take on its maximum values on the boundary C , so that

$$|\phi(x, y) - \psi(x, y)| \leq \max |\phi(\alpha, \beta) - \psi(\alpha, \beta)|.$$

But $\phi(\alpha, \beta) - \psi(\alpha, \beta) = u(\alpha, \beta) - P_{mn}(\alpha, \beta)$ and hence, by (13),

$$|\phi - \psi| \leq \max |u(\alpha, \beta) - P_{mn}(\alpha, \beta)| \leq [(A + \lambda)(2\epsilon)^q/\lambda]^{1/s} = N\epsilon^{q/s},$$

where N is a constant independent of m and n . Thus we can write

$$|u - P_{mn}| \leq M\epsilon^{q/r} + N\epsilon^{q/s} \leq (M + N)\epsilon,$$

a relation which holds uniformly throughout J .

Hence if $u(x, y)$ satisfies the hypothesis of Theorem A so that ϵ can be taken equal to $K\omega(1/m + 1/n)$, it is certain that P_{mn} will converge uniformly in J to the value of u as m and n both become infinite. Hence we can state

THEOREM III. *In problem B in the case $r > 1$, if $u(x, y)$ satisfies the hypothesis of Theorem A, there will exist a positive constant D_3 independent of m and n , and a positive quantity $\epsilon_{mn} = \omega(1/m + 1/n)$ such that the relation*

$$|u - P_{mn}| \leq D_3\epsilon_{mn}$$

holds uniformly throughout J , with $\lim_{m, n \rightarrow \infty} \epsilon_{mn} = 0$.

6. *Problem B. Convergence in the general case $r > 0$.* Let

$$F(x, y) = u(x, y) - p_{mn}(x, y),$$

where p_{mn} is the polynomial of degrees m, n satisfying relations (10) and (11),

$$(10) \quad \dots |\partial^{i+j}(F)/\partial x^i \partial y^j| \leq \epsilon, \quad (i+j=0, 1, 2),$$

$$(11) \quad |\nabla^2(F)| \leq 2\epsilon.$$

Then if Q_{mn} is the polynomial of best approximation to F of degrees m, n , and q is the smaller of the two numbers r, s , we can write

$$\begin{aligned} \gamma &= \iint_J |\nabla^2(F - Q_{mn})|^r dx dy + \lambda \max |F(\alpha, \beta) - Q_{mn}(\alpha, \beta)|^s \\ &\leq \iint_J |\nabla^2 F|^r dx dy + \lambda \max |F(\alpha, \beta)|^s \\ &\leq A(2\epsilon)^r + \lambda \epsilon^s \leq (A + \lambda)(2\epsilon)^q. \end{aligned}$$

Hence, since each term of γ is $\leq \gamma$,

$$(14) \quad \iint_J |\nabla^2(F - Q_{mn})|^r dx dy \leq (A + \lambda)(2\epsilon)^q,$$

$$(15) \quad \max |F(\alpha, \beta) - Q_{mn}(\alpha, \beta)| \leq [(A + \lambda)(2\epsilon)^q/\lambda]^{1/s}.$$

So also, by reason of (15) and (10),

$$(16) \quad |Q_{mn}(\alpha, \beta)| \leq [(A + \lambda)(2\epsilon)^q/\lambda]^{1/s} + \epsilon.$$

Let δ again denote the maximum value of $|\nabla^2 Q_{mn}|$ in J , and let (x_0, y_0) be a point of the region at which $|\nabla^2 Q_{mn}(x_0, y_0)| = \delta$. Then we can show, exactly as in section 4, that a constant E independent of m and n can be found such that

$$(17) \quad \delta \leq E(\bar{m}^4 \epsilon^q)^{1/r}.$$

where \bar{m} is the greater of the two numbers m, n . Moreover Q_{mn} can be written in terms of the Green's function,

$$Q_{mn}(x, y) = \iint_J G(x, y; \xi, \eta) \nabla^2 Q_{mn}(\xi, \eta) d\xi d\eta + \chi(x, y),$$

so that throughout J

$$|Q_{mn}(x, y)| \leq \delta \iint_J |G| d\xi d\eta + \max |\chi(x, y)|.$$

But $\chi(x, y)$ is harmonic in J and therefore acquires its maximum values on the boundary C . Moreover $\chi(\alpha, \beta) = Q_{mn}(\alpha, \beta)$, so that $\max |\chi(x, y)| = \max |Q_{mn}(\alpha, \beta)|$, and therefore, by (16),

$$\max |\chi(x, y)| \leq [(A + \lambda)(2\epsilon)^q/\lambda]^{1/s} + \epsilon \leq M'\epsilon^{q/s},$$

where M' is a constant independent of m and n . By reason of this and (17) it follows that

$$|Q_{mn}| \leq \left[\int_J \int |G| d\xi d\eta \right] E(\bar{m}^4 \epsilon^2)^{1/r} + M' \epsilon^{2/3}.$$

Hence if P_{mn} is the polynomial of best approximation to u of degrees m, n , we can write

$$\begin{aligned} |u - P_{mn}| &= |F - Q_{mn}| \leq |F| + |Q_{mn}| \\ &\leq \epsilon + E(\bar{m}^{4/r} \epsilon^{1/r}) \left[\int_J \int |G| d\xi d\eta \right] + M' \epsilon^{2/3} \leq E' \bar{m}^{4/r} \epsilon, \end{aligned}$$

from which it follows that if u satisfies the hypothesis of Theorem B with the integer k taken $\geq 4/r$, the process converges. Thus we have established

THEOREM IV. *In problem B in the general case $r > 0$, if the curve C is subject to the limitations imposed by the requirements of Markoff's theorem, and if $u(x, y)$ satisfies the hypothesis of Theorem B with k taken $\geq 4/r$, then there will exist a positive quantity $\epsilon_{mn} = \bar{m}^{4/r} (1/m + 1/n)^k \Omega(1/m + 1/n)$, and a positive constant D_4 independent of m and n , such that the relation*

$$|u - P_{mn}| \leq D_4 \epsilon_{mn}$$

holds uniformly throughout J , and furthermore, provided m and n maintain the same order of magnitude,

$$\lim_{m, n \rightarrow \infty} \epsilon_{mn} = 0.$$

MOUNT ALLISON UNIVERSITY,
SACKVILLE, N. B., CANADA.

ON THE INVERSION FORMULA FOR FOURIER-STIELTJES TRANSFORMS IN MORE THAN ONE DIMENSION. II.

By E. K. HAVILAND.

A proof of the Continuity Theorem for multi-dimensional Fourier-Stieltjes transforms based on previous results of the author will be given in the present note. This proof,[†] which for simplicity is given in the case of two dimensions, is believed to be substantially clearer and more direct than the proofs previously given,[‡] the improvement being made possible on the one hand by the use of the Convolution Theorem for Fourier-Stieltjes transforms, first proved generally by the author,[§] and on the other hand by the use of the inversion formula recently proved by the author.[¶] A previous proof^{||} of particular results contained in the complete Convolution Theorem was based on the Continuity Theorem, while the present author's proof of the complete Convolution Theorem is quite independent of it.

We begin by proving a

UNIQUENESS LEMMA.^{††} Let (i) $f(x, y)$ be continuous in $(-\infty < x < +\infty; -\infty < y < +\infty)$,

$$(ii) \quad \iint_S |f(x, y)| \, dx dy < +\infty, \text{ where } S \text{ denotes the entire } (xy)\text{-plane,}$$

$$(iii) \quad \iint_S \exp\{i(sx + ty)\} f(x, y) \, dx dy = 0 \quad \text{for every real } (s, t).$$

Then $f(x, y) \equiv 0$.

[†] The present proof has been developed from a proof of the Continuity Theorem in the one-dimensional case given by A. Wintner in a class on the theory of probability.

[‡] For the one-dimensional case, cf. P. Lévy, *op. cit.*; for the multi-dimensional case, cf. V. Romanovsky, *loc. cit.*, p. 41, and S. Bochner, *loc. cit.*, p. 403. The references are collected at the end of the paper.

[§] Cf. E. K. Haviland, *loc. cit.* II, p. 651, Theorem V.

[¶] Cf. E. K. Haviland, *loc. cit.* III.

Professor C. R. Adams has kindly called my attention to the fact that a statement by B. H. Camp, to the effect that a bounded monotone function is not necessarily of bounded variation, was not intended to refer to functions satisfying *all* the conditions (14) of Hardy to which Camp refers, but that Camp's statement, in its intended sense, is correct, contrary to a remark of the present author in a footnote on p. 95 of the foregoing paper.

^{||} Cf. S. Bochner, *ibid.*; cf. in this connection E. K. Haviland, *loc. cit.* II, p. 626.

^{††} The method of this proof is largely an adaptation of the treatment of a similar problem in one dimension by G. Pólya, *loc. cit.*, pp. 105-106. Cf. also E. K. Haviland, *loc. cit.* II, pp. 638-641.

Proof. Let there be given a rectangle R_1 which may, without loss of generality, be taken to be $(0 \leq x < \xi; 0 \leq y < \eta)$. A function $g_\delta(x, y)$ is defined as follows: $g_\delta(x, y) = 0$ at those points of the rectangle R_2 : $(0 \leq x \leq U; 0 \leq y \leq V)$, where $U > \xi, V > \eta$, which are not in R_1 ; also, $g_\delta(x, y) = 1$ in R_3 : $(\delta \leq x \leq \xi - \delta; \delta \leq y \leq \eta - \delta)$, where $0 < \delta < \text{Min}(\xi/2, \eta/2)$; finally, the value of $g_\delta(x, y)$ at a point (x, y) of $R_1 - R_3$ is given by that point of a truncated pyramid having R_1 as base and R_3 as top whose projection is (x, y) . This function $g_\delta(x, y)$ is extended to the whole plane by prescribing for it the periods U in x and V in y .

As $g_\delta(x, y)$ is continuous everywhere in S , by the two-dimensional Weierstrass trigonometric approximation theorem† there exists a trigonometric polynomial,

$$P_\epsilon(x, y) = \sum_{-M}^M \sum_{-N}^N \alpha_{mn} \exp\{i(2\pi mx/U + 2\pi ny/V)\},$$

such that

$$(1) \quad |g_\delta(x, y) - P_\epsilon(x, y)| < \epsilon$$

for all (x, y) . Setting $s = 2\pi m/U, t = 2\pi n/V$ in (iii), we see that

$$\iint_S \exp\{i(2\pi mx/U + 2\pi ny/V)\} f(x, y) dx dy = 0.$$

Hence $\iint_S P_\epsilon(x, y) f(x, y) dx dy = 0$. We first let $\epsilon \rightarrow 0$ in (1). Since

$$\iint_{S-R} |P_\epsilon(x, y) f(x, y)| dx dy \leq 2 \iint_{S-R} |f(x, y)| dx dy < \eta,$$

where η is arbitrarily small, provided $\epsilon (> 0)$ is sufficiently small and the rectangle R sufficiently large, it follows from the Arzelà-Lebesgue theorem that

$$\iint_S P_\epsilon(x, y) f(x, y) dx dy \rightarrow \iint_S g_\delta(x, y) f(x, y) dx dy, \quad (\epsilon \rightarrow 0),$$

so that the latter integral vanishes.

In the second place, we let $\delta \rightarrow 0$, whereupon $g_\delta(x, y) \rightarrow g(x, y)$, a function equal to one within R_1 and its periodic images and to zero elsewhere. Let the rectangle R_1 now be denoted by R_{10} and let R_{1i} , ($i = 1, 2, 3, \dots$), be periodic images of R_{10} . If R_{2i} be the periodic image of R_2 containing R_{1i} , it follows from (ii) and the inequality $|g_\delta(x, y)| \leq 1$ that

$$(2) \quad \sum_{i=0}^{\infty} \iint_{R_{2i}} g_\delta(x, y) f(x, y) dx dy = \iint_S g_\delta(x, y) f(x, y) dx dy = 0.$$

By again applying the Arzelà-Lebesgue Theorem, we find for every fixed v

† Cf. L. Tonelli, *op. cit.*, p. 494.

$$(3) \quad \sum_{i=0}^{\nu} \int \int_{R_{2i}} g_{\delta}(x, y) f(x, y) dx dy \rightarrow \sum_{i=0}^{\nu} \int \int_{R_{2i}} f(x, y) dx dy, \quad (\delta \rightarrow 0),$$

by the definition of $g(x, y)$. Since this implies that the integral on the right of (3) is zero, on letting $\nu \rightarrow \infty$, it follows that

$$\sum_{i=0}^{\infty} \int \int_{R_{2i}} f(x, y) dx dy = 0.$$

Finally, we let $U \rightarrow +\infty$, $V \rightarrow +\infty$, whereupon we obtain, in view of the absolute convergence of the foregoing series and of the continuity of $f(x, y)$ in R_{10} ,

$$\int_0^{\xi} dx \int_0^{\eta} f(x, y) dy = \int \int_{R_{20}} f(x, y) dx dy = 0.$$

Since we may choose (ξ, η) arbitrarily and since $f(x, y)$ is continuous, we may differentiate the latter integral with respect to ξ and η , obtaining $f(\xi, \eta) \equiv 0$, q. e. d. It is to be noted that the hypothesis (i) was used only in the final step of the proof of the lemma.

We are now in a position to prove the

CONTINUITY THEOREM FOR FOURIER-STIELTJES TRANSFORMS.† If $\{\phi_n\}$ be a sequence of distribution functions and $\{\Lambda(s, t; \phi_n)\}$ the sequence of corresponding Fourier-Stieltjes transforms, then a necessary and sufficient condition that the sequence $\{\phi_n\}$ should converge to a distribution function ϕ is that the sequence $\{\Lambda(s, t; \phi_n)\}$ converges to a function $h(s, t)$ uniformly in every finite region of the (s, t) -plane. Furthermore, $h(s, t) = \Lambda(s, t; \phi)$.

Proof. We first prove the sufficiency of the condition, noting that as a consequence of our hypothesis $h(s, t)$ is continuous at every point of the (s, t) -plane and $|h(s, t)| \leq 1$.

Let $\gamma(E)$ be the two-dimensional Gaussian distribution function; i. e., to $\gamma(E)$ corresponds ‡ the point function

$$G(x, y) = (2\pi)^{-1} \int_{-\infty}^x \int_{-\infty}^y \exp\{-(\xi^2 + \eta^2)/2\} d\xi d\eta.$$

As the Fourier-Stieltjes transform of $\gamma(E)$ may be regarded as an iterated integral, its value may be computed from the known result in the case of one dimension to be

$$(4) \quad \Lambda(s, t; \gamma) = \exp\{-(s^2 + t^2)/2\}.$$

We then set

† For definition of terms occurring in this theorem, cf., e. g., E. K. Haviland, *loc. cit.* II, pp. 627-628.

‡ Cf. J. Radon, *loc. cit.*, p. 1304; E. K. Haviland, *loc. cit.* II, p. 627.

$$(5) \quad L(s, t) = h(s, t) \Lambda(s, t; \gamma).$$

It is a continuous function of (s, t) and

$$(6) \quad |L(s, t)| \leq |\Lambda(s, t; \gamma)| = \Lambda(s, t; \gamma).$$

Let $\{\phi_{m_n}\}$ be a convergent subsequence of $\{\phi_n\}$ and $\tau = \tau(E)$ be its limit. $\tau(E)$ is monotone by the Compactness Theorem of Radon † and $0 \leq \tau(E) \leq 1$ for all E . We next put ‡

$$(7) \quad \rho_n = \phi_{m_n} * \gamma$$

and

$$(7^a) \quad \rho = \tau * \gamma.$$

Since §

$\Lambda(s, t; \rho_n) = \Lambda(s, t; \phi_{m_n}) \cdot \Lambda(s, t; \gamma)$ and $|\Lambda(s, t; \phi_{m_n})| \leq \int_S \int_S d_{xy} \phi_{m_n}(E) = 1$, it follows that

$$(8) \quad |\Lambda(s, t; \rho_n)| \leq \Lambda(s, t; \gamma),$$

uniformly with respect to n . Similarly,

$$|\Lambda(s, t; \rho)| \leq \Lambda(s, t; \gamma).$$

ρ_n and ρ are both continuous by virtue of the addition rule of line spectra. We proceed to show that, as $n \rightarrow \infty$, $\rho_n(R) \rightarrow \rho(R)$ for every rectangle R . Not only does

$$\rho_n(R) = \int_S \int_S \phi_{m_n}(R - P_{xy}) \exp\{-(x^2 + y^2)/2\} dx dy$$

exist for every R , due to the continuity of γ , but the integrand of ρ_n has a bounded and absolutely integrable majorant independent of n . Also, $\phi_{m_n}(E) \rightarrow \tau(E)$ on all non-singular lines of the latter as $n \rightarrow \infty$. Then, by the Arzelà-Lebesgue Theorem, as $n \rightarrow \infty$,

$$(10) \quad \rho_n(R) \rightarrow \int_S \int_S \tau(R - P_{xy}) \exp\{-(x^2 + y^2)/2\} dx dy = \tau * \gamma = \rho.$$

Since ρ_n and ρ have no singular rectangles, we obtain by the inversion formula for Fourier-Stieltjes transforms ¶

† Cf. E. K. Haviland, *loc. cit.* I, p. 551.

‡ $\psi_1 * \psi_2$ denotes the symbolical product (Faltung or convolution) of ψ_1 and ψ_2 . It is sufficient for its existence that ψ_1 and ψ_2 be monotone bounded functions, in which case the addition rule of spectra also holds. Cf. E. K. Haviland, *loc. cit.* II, p. 654.

§ This follows from the Convolution Theorem for Fourier-Stieltjes transforms. Cf. E. K. Haviland, *loc. cit.* II, p. 651, Theorem V. It is important for what follows to note that the theorem holds for any two arbitrary monotone bounded functions.

¶ Cf. E. K. Haviland, *loc. cit.* III, p. 99, equation (8).

$$(11) \quad - (2\pi)^2 \rho_n(R) \\ = \int \int_S (st)^{-1} \Lambda(s, t; \rho_n) [e^{-is\xi} - 1] [e^{-it\eta} - 1] e^{-i(us+vt)} ds dt,$$

$$(12) \quad - (2\pi)^2 \rho(R) \\ = \int \int_S (st)^{-1} \Lambda(s, t; \rho) [e^{-is\xi} - 1] [e^{-it\eta} - 1] e^{-i(us+vt)} ds dt.$$

It is not necessary to use Cauchy principal values in these equations, as both $\Lambda(s, t; \rho_n)$ and $\Lambda(s, t; \rho)$ possess absolutely integrable majorants in virtue of (8), (9) and (4). It follows from (10) that, as $n \rightarrow \infty$, the left-hand side of (11) approaches the left-hand side of (12). In consequence, the right-hand side of (11) must approach the right-hand side of (12).

Now from (4), (6) and (8), together with the fact that $[e^{-is\xi} - 1]/s$, $[e^{-it\eta} - 1]/t$ are uniformly bounded for all (s, t) , it follows that the Arzelà-Lebesgue Theorem may be applied to the right-hand side of (11), so that, as $n \rightarrow \infty$,

$$(13) \quad - (2\pi)^2 \rho_n(R) \rightarrow \int \int_S (st)^{-1} L(s, t) [e^{-is\xi} - 1] [e^{-it\eta} - 1] e^{-i(us+vt)} ds dt$$

in virtue of (5), (7) and the Convolution Theorem for Fourier-Stieltjes transforms. Hence, by the last remark of the preceding paragraph,

$$(14) \quad \int \int_S f(s, t) (st)^{-1} [e^{-is\xi} - 1] [e^{-it\eta} - 1] e^{-i(us+vt)} ds dt = 0,$$

where $f(s, t) = L(s, t) - \Lambda(s, t; \rho)$, so $|f(s, t)| \leq 2\Lambda(s, t; \gamma)$, which implies the absolute integrability of the integrand in (14). Then we may differentiate with respect to ξ and η beneath the integral sign in (14), obtaining

$$(15) \quad \int \int_S f(s, t) \exp\{-i[s(\xi + u) + t(\eta + v)]\} ds dt = 0.$$

From (5) and from the definition of $f(s, t)$, together with the fact that (15) holds for all $(\xi + u)$, $(\eta + v)$, it follows that $f(s, t)$ satisfies the conditions of the Uniqueness Lemma, so $f(s, t) \equiv 0$, or $L(s, t) \equiv \Lambda(s, t; \rho)$, or by (5) and (7), as $\Lambda(s, t; \gamma) \neq 0$,

$$(16) \quad h(s, t) = \Lambda(s, t; \tau).$$

Consequently, $\Lambda(s, t; \tau)$ does not depend on the special choice of the subsequence $\{\phi_{m_n}\}$ and as (by the inversion formula) τ is determined up to its singular lines by its Fourier-Stieltjes transform, it follows that τ does not depend on the special choice of $\{\phi_{m_n}\}$. This implies that $\{\phi_n\}$ is convergent, for otherwise it would be possible to select from $\{\phi_n\}$ two subsequences converging to essentially distinct limits, say τ_1 and τ_2 . As, however, τ_1 and τ_2

have the same Fourier-Stieltjes transforms, this leads to a contradiction. Finally, if we set $s = t = 0$ in (16), we see that $\Lambda(0, 0; \tau) = 1$, so τ is indeed a distribution function. As τ may thus be taken as ϕ , this completes the proof of the first half of the theorem.

To prove the second half of the theorem, we set $\exp\{i(sx + ty)\} = g(s, t; x, y)$ and let J be a non-singular square of ϕ so large that

$$(17) \quad 0 \leq \phi(S - J) < \epsilon.$$

Then let N'_ϵ be chosen so large that $|\phi_n(S - J) - \phi(S - J)| < \epsilon$ for all $n \geq N'_\epsilon$. It follows that for all such n

$$(18) \quad 0 \leq \phi_n(S - J) < 2\epsilon.$$

We next take a division of J : ($-M \leq x \leq M$; $-M \leq y \leq M$) by drawing parallels to the axes, these parallels being non-singular lines of ϕ and dividing J into a finite number, m , of rectangles R_k whose greatest diameter is δ_m , $\lim_{m \rightarrow \infty} \delta_m = 0$. By choosing $\delta_m < \delta = \delta(\epsilon)$, we can make

$$|g(s, t; x_k, y_k) - g(s, t; x'_k, y'_k)| < \epsilon,$$

where (x_k, y_k) , (x'_k, y'_k) are any two points of R_k , and $\delta(\epsilon)$ is independent of (s, t) in an arbitrarily fixed closed rectangle Σ of the st -plane. Then if $\delta_m < \delta$, we have †

$$(19) \quad \left| \sum_{k=1}^m g(s, t; x_k, y_k) \phi_n(R_k) - \iint_S g(s, t; x, y) d_{xy} \phi_n(E) \right| < \epsilon \iint_S d_{xy} \phi_n(E) = \epsilon,$$

and similarly

$$(20) \quad \left| \sum_{k=1}^m g(s, t; x_k, y_k) \phi(R_k) - \iint_S g(s, t; x, y) d_{xy} \phi(E) \right| < \epsilon.$$

But m being fixed when δ is chosen and the m rectangles R_k being non-singular rectangles of ϕ ,

$$(21) \quad \left| \sum_{k=1}^m g(s, t; x_k, y_k) \phi(R_k) - \sum_{k=1}^m g(s, t; x_k, y_k) \phi_n(R_k) \right| \leq \sum_{k=1}^m |\phi(R_k) - \phi_n(R_k)| < \epsilon,$$

provided $n \geq N''_\epsilon$. Hence if $N_\epsilon = \text{Max}(N'_\epsilon, N''_\epsilon)$, it follows from (17), (18), (19), (20), (21) that

$$\begin{aligned} & \left| \iint_S g(s, t; x, y) d_{xy} \phi(E) - \iint_S g(s, t; x, y) d_{xy} \phi_n(E) \right| \\ &= |\Lambda(s, t; \phi) - \Lambda(s, t; \phi_n)| < 6\epsilon, \end{aligned}$$

† Cf. J. Radon, *loc. cit.*, p. 1324, equation (14).

provided $n \geq N_\epsilon$, where N_ϵ is independent of (s, t) in the arbitrarily fixed rectangle Σ , q. e. d.

COROLLARY. If, as $n \rightarrow \infty$, the sequence of Fourier-Stieltjes transforms $\{\Lambda(s, t; \phi_n)\}$ converges in the whole (s, t) -plane to a continuous function $h(s, t)$ then the convergence is uniform in every finite region of the (s, t) -plane.

Proof. Bochner has shown [†] that the convergence of $\{\Lambda(s, t; \phi_n)\}$ to a continuous function $h(s, t)$ is a sufficient condition for the convergence of $\{\phi_n\}$ to a distribution function ϕ , while we have shown that the uniform convergence of the sequence of Fourier-Stieltjes Transforms in every finite region of the (s, t) -plane is both a necessary and a sufficient condition for the essential convergence of $\{\phi_n\}$ to a distribution function ϕ . Thus Bochner's statement of the Continuity Theorem [‡] is in reality no more general than the usual § formulation of the theorem.

REFERENCES.

- S. Bochner, "Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse," *Mathematische Annalen*, vol. 108 (1933), pp. 378-410.
- E. K. Haviland, I: "On statistical methods in the theory of almost-periodic functions," *Proceedings of the National Academy of Sciences*, vol. 19 (1933), pp. 549-555.
- II: "On the theory of absolutely additive distribution functions," *American Journal of Mathematics*, vol. 56 (1934), pp. 625-658.
- III: "On the inversion formula for Fourier-Stieltjes transforms in more than one dimension," *ibid.*, vol. 57 (1935), pp. 94-100.
- P. Lévy, *Calcul des probabilités* (Paris, 1925).
- G. Pólya, "Herleitung des Gausschen Fehlergesetzes aus einer Funktionalgleichung," *Mathematische Zeitschrift*, vol. 18 (1923), pp. 96-108.
- J. Radon, "Theorie und Anwendungen der absolut additiven Mengenfunktionen," *Sitzungsberichte der mathematischen-naturwissenschaftlichen Klasse der Kaiserl. Akademie zu Wien*, vol. 122 (1913), pp. 1295-1438.
- V. Romanovsky, "Sur un théorème limite du calcul des probabilités," *Recueil Mathématique de la Société Mathématique de Moscou*, vol. 36 (1929), pp. 36-64.
- L. Tonelli, *Serie Trigonometriche* (Bologna, 1928).

THE JOHNS HOPKINS UNIVERSITY.

[†] Cf. S. Bochner, *loc. cit.*, p. 403, Theorem 17. In this connection, it may be noted that our proof for our sufficient condition may be used without modification to prove Bochner's Theorems 16 and 17, save that in the former case the integrals must be considered as Lebesgue integrals, so that from the differentiation of our equation (14) we may conclude only that $f(s, t) = 0$ almost everywhere.

[‡] Cf. J. Radon, *loc. cit.*, p. 1324, equation (14).

§ Cf. P. Lévy, *op. cit.*

ISOLATED CRITICAL POINTS.

By ARTHUR B. BROWN.

The object of this note is to replace an incomplete proof of an earlier paper* by a proof using the methods of that paper. Professor Marston Morse, originator of the general theory of critical points, who pointed out to the writer that in the proof of Lemma 14 of BI it is not shown that a deformation is determined, has published results of which this Lemma 14 is a corollary.† The treatment‡ to follow is of different nature from the treatment of the point in question by Morse.

Proof of Lemma 14. We subdivide the complex D (defined on page 265 of BI) regularly at least once till the D -neighborhoods,§ say \mathcal{N}_a , of the centers P of the spheres S , with boundaries, are interior to the spheres S . If we remove the points P from \mathcal{N}_a , then the remainder, \mathcal{N}'_a , of \mathcal{N}_a is covered by a field \mathcal{F} of curves, each curve joining a point P to a point of $W = \bar{\mathcal{N}}_a - \mathcal{N}_a$, as follows easily from the structure of a simplicial complex. Let B'' be the set defined like B' , but for smaller spheres, say S_2 , so that any point of W is outside all the spheres S_2 . If we shrink \mathcal{N}'_a down onto W by use of the field \mathcal{F} , then the resulting deformation, say (D_1) , carries D' over itself into a subset of B'' . Points outside $\bar{\mathcal{N}}_a$ remain fixed under (D_1) .

Let $[\Sigma]$ be a set of spheres slightly larger than S , concentric with the latter and satisfying the same conditions. Choose $\epsilon > 0$ so small that the

* A. B. Brown, "Relations between the critical points of a real analytic function of n independent variables," *American Journal of Mathematics*, vol. 52 (1930), pp. 251-270. We refer to this paper as BI. Cf. footnote 3 of the writer's paper, "Critical sets of an arbitrary real analytic function of n variables," *Annals of Mathematics*, vol. 32 (1931), pp. 512-520.

† Marston Morse, "The critical points of a function of n variables," *Transactions of the American Mathematical Society*, vol. 33 (1931), pp. 72-91 (Morse I). Lemma 14 of BI is a corollary of Theorem 9, page 84, of Morse I. See also Theorem 5.I, page 156, of Marston Morse, "The calculus of variations in the large," *American Mathematical Society Colloquium Publications*, vol. 18, New York, 1934 (Morse II). For other papers on critical points see bibliography of Morse II.

‡ The writer does not know whether the questionable statement in the "proof" of Lemma 14 in BI is or is not true. Shortly before the appearance of Morse's *Colloquium* the writer, having momentarily forgotten that Lemma 14 follows from results in Morse I, devised the present proof.

§ That is, sets of all cells of D having a vertex at any of the points P . For notations in topology see S. Lefschetz, "Topology," *American Mathematical Society Colloquium Publications*, vol. 12, New York, 1930. That complexes in the sense of analysis situs are at hand is proved by B. O. Koopman and A. B. Brown, *Transactions of the American Mathematical Society*, vol. 34 (1932), pp. 231-251; also by S. Lefschetz and J. H. C. Whitehead, *ibid.*, vol. 35 (1933), pp. 510-517. The fact that complexes are present was used in BI.

trajectories τ (§ 9 of BI)* exist between and on the pairs of spheres Σ and S_2 , at points where $c - \epsilon \leq f \leq c$. Recall that the parameter for the trajectories τ is the distance r from P , in any sphere Σ . Let a_{s_2} and a_σ denote the radii of S_2 and Σ respectively, and M the minimum distance from the locus $f \leq c - \epsilon/2$ to the locus $f = c$, between or on the pairs of spheres S_2 and Σ . Consider the transformation which acts only upon the points Q of S_2 satisfying $c - \epsilon \leq f \leq c$ sending each such point into a point Q' on the same trajectory τ , and determined by

$$(1) \quad r' = r + \frac{(f - c + \epsilon)}{\epsilon} \cdot (a_\sigma - a_{s_2}).$$

We now determine a deformation (D_2) which keeps fixed all points except those on the trajectories τ between the pairs of points Q and Q' . The deformation causes each of the trajectories QQ' to shrink down to the point Q' , and is defined in an obvious way in terms of r .

Since $f = \text{constant}$ on any trajectory τ , it follows from (1) that B'' is carried by (D_2) into a set whose points within distance $\frac{1}{2}(a_\sigma - a_{s_2})$ from S_2 satisfy $f \leq c - \epsilon/2$, and hence are distant at least M from the part of the locus $f = c$ between or on the spheres S_2 and Σ . Hence we can follow (D_2) by a deformation along radial lines through P in each sphere Σ , affecting only points within distance $U_{s_2} = \min. [M, \frac{1}{2}(a_\sigma - a_{s_2})]$ of S_2 , so that, as a result of the two deformations, locus B'' is deformed over itself into a subset of the corresponding locus for spheres of radius $a_{s_2} + U_{s_2}$. It is then clear that a finite number of such steps will deform B'' over itself into a subset of B' , with B' remaining on B' during the entire resulting deformation (D_3) .

If now we perform (D_1) and then (D_3) , it is seen that the resulting deformation (D_4) carries D' into a subset of B' , while keeping B' on B' . From Theorem 2, page 252, of BI, it follows that B' and D' have the same Betti numbers, and Lemma 14 is proved.†

COLUMBIA UNIVERSITY.

* In the more general case treated by Morse, the trajectories τ become the (ϕf) -trajectories (Morse I, page 80; Morse II, page 153). The $(f\phi)$ -trajectories of Morse's treatment do not appear in BI.

† We wish also to point out that on page 261 of BI, the definition of configuration is not given properly. In lines 7 to 2 from the bottom, "when $n - s \dots$ (ordinary points)" should be replaced by "when, after a non-singular linear transformation, $n - s$ of the variables are the values of the dependent variables defined by the vanishing of $n - s$ algebroid functions (pseudopolynomials), where the other s variables, say ξ_1, \dots, ξ_s , are the independent variables for each algebroid function. These values are analytic at points where the discriminants of the algebroid functions are not zero (ordinary points)". On page 262, line 4, " x_1, \dots, x_s . Therefore if the" is replaced by " ξ_1, \dots, ξ_s . Therefore if a". In line 6, "discriminant" is replaced by "discriminants, separately and severally". In line 12, delete "variables x_1, x_2, \dots, x_n as".

CYCLOTOMY, HIGHER CONGRUENCES, AND WARING'S PROBLEM.

By L. E. DICKSON.

1. *Introduction.* This memoir does not presuppose any knowledge of the subjects treated. The outstanding Waring problem is to find s such that every large integer is a sum of s positive integral values of a given polynomial. An account of its recent solution is given in § 29. One step is the proof that every integer is congruent to a sum of s values of the polynomial with respect to every prime modulus p and certain powers of p . The proof employs the number N of solutions of $x^e + y^e + 1 \equiv 0 \pmod{p = ef + 1}$.

By far the most effective method of finding N is that of cyclotomy, which yields also the number of solutions of any trinomial congruence involving three e -th powers multiplied by any integers.

The periods can be expressed by radicals in terms of certain resolvent functions. But this algebraic side of cyclotomy has little practical application to our problem to find the e^2 cyclotomic constants (k, h) , which are coefficients in the product of two periods expressed linearly in terms of the periods.

Unfortunately the latter problem has been solved heretofore only when $e \leq 5$ and then the problem is so simple † that there arise none of the difficulties for $e \geq 6$.

While $e = 6$ had been treated, the solution involved the six numbers A, \dots, F in the decompositions

$$p = A^2 + 3B^2, \quad 4p = L^2 + 27M^2, \quad 4p = E^2 + 3F^2,$$

whereas the true solution (§ 17) of the problem involves only A and B . A similarly perfect solution is obtained for the new cases $e = 8, 10, 12$. The odd values of e are not needed in Part 2.

Our methods serve for further values of e . But the results must be postponed to later papers.

* We need a formula for N which implies that N will exceed any given number when p exceeds an obtainable limit. In *Journal de Mathématiques*, vol. 2 (1837), pp. 253-292, V. A. Lebesgue found that N is congruent modulo p to a long sum of binomial coefficients. But this result does not yield the needed property.

† Except for the proof when $e = 5$ that the pair of Diophantine equations have essentially a unique solution.

PART I. CYCLOTOMY, HIGHER CONGRUENCES.

2. *The periods.* Let g be a primitive root of a prime p . Let e be a divisor of $p-1$ and write $p-1=ef$. Let R be any (imaginary) root $\neq 1$ of $x^e=1$. The sums

$$(1) \quad \eta_k = \sum_{t=0}^{f-1} R^{g^{et+k}} \quad (k=0, 1, \dots, e-1)$$

are called *periods*. For example, if $p=7$, $e=3$, then $f=2$ and 3 is a value of g . Since $g^2 \equiv 2$, $g^3 \equiv 6 \pmod{7}$, the periods (1) are

$$\eta_0 = R + R^6, \quad \eta_1 = R^3 + R^4, \quad \eta_2 = R^5 + R^5.$$

Let s be the summation index for η_0 . For a fixed s , we may replace t in (1) by $t+s$, which ranges with t over a complete set of residues modulo f . Hence

$$(2) \quad \eta_0 \eta_k = \sum_{s=0}^{f-1} \sum_{t=0}^{f-1} R^{g^{esN}}, \quad (N=1+g^{et+k}).$$

First, let $N \equiv 0 \pmod{p}$. Since $0 \leq et+k \leq ef-1 \leq p-2$,

$$et+k = \frac{1}{2}(p-1).$$

If f is even, k is divisible by e , whence $k=0$, $t=f/2$. But if f is odd, k is divisible by $e/2$, while $k \neq 0$ since $ef/2$ is not divisible by e , whence $k=e/2$, $t=(f-1)/2$. Make the definition

$$(3) \quad \begin{aligned} n_k &= 1 \text{ if } f \text{ is even and } k=0, \text{ or if } f \text{ is odd and } k=e/2; \\ n_k &= 0 \text{ in all remaining cases.} \end{aligned}$$

Hence $N \equiv 0 \pmod{p}$ holds for exactly n_k values of t , and the corresponding part of (2) is $f n_k$.

Second, let N be prime to p , whence N is congruent to a power of the primitive root g :

$$(4) \quad 1 + g^{et+k} \equiv g^{ez+h} \pmod{p},$$

where $0 \leq h \leq e-1$, $0 \leq z \leq f-1$. When h (as well as k) is fixed, let

$$(5) \quad \begin{aligned} (k, h) &\text{ be the number of sets of values of } t \text{ and } z, \\ &\text{each chosen from } 0, 1, \dots, f-1, \text{ for which (4) holds.} \end{aligned}$$

Hence (k, h) is unaltered if we increase (or decrease) either k or h by any multiple of e . For fixed values of t and z satisfying (4), the corresponding part of (2) is

$$\sum_{s=0}^{f-1} R^{g^{e(s+z)+h}} = \sum_{s=0}^{f-1} R^{g^{es+h}} = \eta_h,$$

since $s+z$ ranges with s over a complete set of residues modulo f . This completes the proof of

$$(6) \quad \eta_0 \eta_k = \sum_{h=0}^{e-1} (k, h) \eta_h + f n_k \quad (k = 0, 1, \dots, e-1).$$

Replace R by R^{g^m} . Then η_k becomes η_{k+m} , in which we may reduce subscripts of η modulo e . Hence

$$(7) \quad \eta_m \eta_{m+k} = \sum_{h=0}^{e-1} (k, h) \eta_{m+h} + f n_k.$$

3. *The period equation.* Since the e periods (1) contain without duplication R, \dots, R^{p-1} , whose sum is -1 ,

$$(8) \quad 1 + \eta_0 + \eta_1 + \dots + \eta_{e-1} = 0.$$

Employ also (6) for $k = 1, \dots, e-1$. Regard η_0 as a constant. We have e linear homogeneous equations in $1, \eta_1, \dots, \eta_{e-1}$. Hence

$$(9) \quad \begin{vmatrix} 1 + \eta_0 & 1 & \dots & 1 \\ f n_1 + (1, 0) \eta_0 & (1, 1) - \eta_0 & \dots & (1, e-1) \\ \vdots & \vdots & \ddots & \vdots \\ f n_{e-1} + (e-1, 0) \eta_0 & (e-1, 1) & \dots & (e-1, e-1) - \eta_0 \end{vmatrix} = 0,$$

which is the period equation satisfied by η_0 and also by every η_k .

4. *Auxiliary congruence.* The number of sets of values of t and z , each chosen from $0, 1, \dots, f-1$, which satisfy

$$(10) \quad 1 + g^{et+k} + g^{ez+h} \equiv 0 \pmod{p}$$

will be denoted by $\{k, h\} = \{h, k\}$. Evidently $\{k, h\}$ is unaltered if we increase k and h by multiples of e . Multiply (10) by the reciprocal of its second term; we get

$$1 + g^{e(-t)-k} + g^{e(z-t)+h-k} \equiv 0 \pmod{p}.$$

Since $-t$ and $z-t$ uniquely determine t and z modulo f ,

$$(11) \quad \{-k, h-k\} = \{k, h\}.$$

We may express $\{k, h\}$ in terms of our former (i, j) . First, let f be even. Then

$$p-1 = 2e \cdot f/2, \quad -1 \equiv g^{(p-1)/2} = g^{e \cdot f/2} \pmod{p}.$$

Thus (10) may be written as

$$1 + g^{et+k} \equiv g^{e(z+f/2)+h} \pmod{p}.$$

Comparison with (4) gives

$$(12) \quad \{k, h\} = (k, h), \quad f \text{ even.}$$

For f odd, (10) may be written as

$$(13) \quad 1 + g^{et+k} \equiv g^m \pmod{p}, \quad m = e[z + \tfrac{1}{2}(f-1)] + h + \tfrac{1}{2}e, \\ \{k, h\} = (k, h + \tfrac{1}{2}e), \quad f \text{ odd.}$$

From $\{k, h\} = \{h, k\}$, (11), (12), (13), we get

$$(14) \quad (k, h) = (h, k), \quad (e-k, h-k) = (k, h), \quad f \text{ even,}$$

$$(15) \quad (k, h) = (h + \tfrac{1}{2}e, k + \tfrac{1}{2}e), \quad (e-k, h-k) = (k, h), \quad f \text{ odd.}$$

By (12) and (13), the systems (14) and (15) are permuted when

$$(16) \quad (k, h) \text{ corresponds to } (k, h + \tfrac{1}{2}e).$$

5. *Linear relations.* The sum (2) involves f^2 powers of R . In (6) the number of powers of R (including 1) is $\Sigma(k, h)f + fn_k$. Cancelling f , we get

$$(17) \quad \sum_{h=0}^{e-1} (k, h) = f - n_k \quad (k = 0, 1, \dots, e-1).$$

It may be verified by (14) and (15) that we may discard as redundant those relations (17) in which $k > e/2$ if e is even, but $k > (e-1)/2$ if e is odd and hence f even.

6. *Case $e = 2$.* We employ (3), (14)-(17). For f even,

$$(18) \quad (0, 0) + (0, 1) = f - 1, \quad (1, 0) + (1, 1) = f, \\ (1, 1) = (1, 0) = (0, 1) = f/2, \quad (0, 0) = \tfrac{1}{2}f - 1.$$

For f odd,

$$(19) \quad (0, 0) + (0, 1) = f, \quad (1, 0) + (1, 1) = f - 1, \\ (0, 0) = (1, 1) = (1, 0) = (f-1)/2, \quad (0, 1) = (f+1)/2.$$

Hence for every f , the (ij) are uniquely determined by $p = 2f + 1$. The period equation (9) is $\eta^2 + \eta + c = 0$, where $c = fn_1 - (1, 1)$, $c = -\frac{1}{4}(p-1)$ if f even, $c = \frac{1}{4}(p+1)$ if f odd.

7. When $e \geq 3$, (14)-(17) do not determine the (k, h) , but must be supplemented by relations obtained by the following advanced theory. By (7),

$$\sum_{j=0}^{e-1} \eta_j \eta_{j+k} = \sum_{j=0}^{e-1} \sum_{h=0}^{e-1} (k, h) \eta_{j+h} + ef n_k.$$

For a fixed h we may replace j by $j - h$; the double sum becomes

$$\Sigma_j [\Sigma_h (k, h)] \eta_j = \Sigma_j (f - n_k) \eta_j = - (f - n_k),$$

by (17) and (18). Also

$$(20) \quad \begin{aligned} ef n_k - (f - n_k) &= (ef + 1) n_k - f = p n_k - f, \\ \sum_{j=0}^{e-1} \eta_j \eta_{j+k} &= p n_k - f \quad (k = 0, \dots, f-1). \end{aligned}$$

8. *Jacobi's functions.* Let α be any root $\neq 1$ of $\alpha^{p-1} = 1$. Write

$$(21) \quad F(\alpha) = \sum_{k=0}^{p-2} \alpha^k R^{\sigma^k}.$$

Usually we employ a special case of this function (21) due to Jacobi. Let $p = ef + 1$ and let β be a primitive e -th root of unity. In (21) take $\alpha = \beta^n$, write $k = et + j$ and employ (1). Thus

$$(22) \quad F(\beta^m) = \sum_{j=0}^{e-1} \beta^{mj} \eta_j.$$

Consider its product by $F(\beta^n)$. For j fixed, the summation index in $F(\beta^n)$ may be taken to be $j + k$, which ranges with k over a complete set of residues modulo e . Hence

$$(23) \quad \begin{aligned} F(\beta^m) F(\beta^n) &= \sum_{k=0}^{e-1} \sum_{j=0}^{e-1} \beta^{mj} \beta^{n(j+k)} \eta_j \eta_{j+k}, \\ F(\beta^m) F(\beta^n) &= \sum_{k=0}^{e-1} \beta^{nk} M_k, \quad M_k = \sum_{j=0}^{e-1} \beta^{(m+n)j} \eta_j \eta_{j+k}. \end{aligned}$$

First, let $m = -n$, where n is not a multiple of e . Thus M_k has the value (20). Since the sum of the n -th powers of the roots $1, \beta, \dots, \beta^{e-1}$ of $x^e = 1$ is zero,

$$(24) \quad \sum_{j=0}^{e-1} \beta^{nj} = 0.$$

Transpose the term given by $k = 0$ or $k = e/2$ according as f is even or odd, and apply (3). Note that if f is odd, e is even and $\beta^{e/2} = -1$. Hence

$$(25) \quad F(\beta^n) F(\beta^{-n}) = (-1)^{nf} p, \quad n \text{ not divisible by } e.$$

Second, let no one of n , m , $n + m$ be divisible by e . Write

$$N_k = \sum_{j=0}^{e-1} \beta^{(m+n)j} \sum_{h=0}^{e-1} (k, h) \eta_{j+h}.$$

By (7) with m replaced by j and (24),

$$M_k - N_k = f n_k \sum_{j=0}^{e-1} \beta^{(m+n)j} = 0.$$

Evidently N_k is the product of

$$F(\beta^{m+n}) = \sum_{j=0}^{e-1} \beta^{(m+n)(j+h)} \eta_{j+h}, \quad \sum_{h=0}^{e-1} \beta^{-(m+n)h} (k, h).$$

Since the first sum is independent of k ,

$$(26) \quad \frac{F(\beta^m) F(\beta^n)}{F(\beta^{m+n})} = \sum_{k=0}^{e-1} \beta^{nk} \sum_{h=0}^{e-1} \beta^{-(m+n)h} (k, h) = R(m, n)$$

(no one of m , n , $m + n$ divisible by e).

We may shorten the computation of $R(m, n)$ by combining its terms in pairs. By the second relation in (14) or (15), $(e - j, h) = (j, j + h)$. Hence the part of (26) given by $k = e - j$ with $j \geq 1$ is equal to

$$\beta^{n(e-j)} \sum_{h=0}^{e-1} \beta^{-(m+n)h} (j, j + h).$$

We may replace the index h by $h - j$ and get

$$\beta^{mj} \sum_{h=0}^{e-1} \beta^{-(m+n)h} (j, h).$$

If $j < e/2$, we may combine this with the new term of R given by $k = j$. Write $E = e/2$ if e is even, $E = (e - 1)/2$ if e is odd. Thus

$$(27) \quad R(m, n) = \sum_{j=1}^E (\beta^{mj} + \beta^{nj}) \sum_{h=0}^{e-1} \beta^{-(m+n)h} (j, h) + \sum_{h=0}^{e-1} \beta^{-(m+n)h} (0, h),$$

when e is odd. But for e even, (27) holds only when in the term given by $j = E$ we replace $\beta^{mj} + \beta^{nj}$ by β^{mE} if $m \equiv n \pmod{2}$, but by zero if $m \not\equiv n \pmod{2}$.

Employ (25) also with n replaced by m and by $m + n$. Then (26) gives

$$(28) \quad R(m, n) R(-m, -n) = p, \text{ none of } m, n, m + n \text{ divisible by } e;$$

$$(29) \quad R(-m, -n) \text{ is derived from } R(m, n) \text{ by replacing } \beta \text{ by } \beta^{-1}.$$

9. *Case $e = 3$.* For a prime $p = 3f + 1$, f is even. By (14),

$$(30) \quad (10) = (01), (11) = (02), (20) = (02), (21) = (12), (22) = (01).$$

Hence the nine (ij) reduce to $(00), (01), (02), (12)$. By (17) and (27),

$$(31) \quad (00) + (01) + (02) = f - 1, \quad (01) + (02) + (12) = f,$$

$$(32) \quad R(1, 1) = u + 3\beta M, u = (00) + 2(12) - 3(02), M = (01) - (02).$$

By (28) and (29),

$$4p = 4(u + 3\beta M)(u + 3\beta^2 M) = (2u - 3M)^2 + 27M^2.$$

Multiply equations (31) by 2 and 5, and subtract. Thus

$$2(00) - 3(01) - 3(02) - 5(12) = -3f - 2.$$

Hence $2u - 3M = L = 9(12) - p - 1$. Thus

$$(33) \quad 4p = L^2 + 27M^2, \quad L \equiv 1 \pmod{3},$$

$$(34) \quad 9(12) = p + 1 + L, \quad 9(00) = p - 8 + L,$$

$$(35) \quad 18(01) = 2p - 4 - L + 9M, \quad 18(02) = 2p - 4 - L - 9M.$$

Hence by (30) all nine (ij) are expressed in terms of p, L, M . By the theory of binary quadratic forms, L^2 and M^2 are uniquely determined by (33). The sign of L has been chosen so that congruence (33) holds. But the sign of M depends on the primitive root g employed; see below (93).

10. Higher congruences.

THEOREM 1. *If no c_i is divisible by the prime $p = ef + 1$, the number of solutions x_1, \dots, x_n all prime to p of*

$$(36) \quad \sum_{i=1}^n c_i x_i^e \equiv d \pmod{p}$$

is e^n times the number of sets of values of z_1, \dots, z_n , each chosen from $0, 1, \dots, f - 1$, which satisfy

$$(37) \quad \sum_{i=1}^n g^{ez_i + a_i} \equiv d \pmod{p},$$

where g is a primitive root of p and $c_i \equiv g^{a_i} \pmod{p}$.

We may write $x_i \equiv g^{y_i} \pmod{p}$, $0 \leq y_i \leq p - 2$. Divide y_i by f . Then

$y_i = q_i f + z_i$, $0 \leq z_i \leq f-1$, $0 \leq q_i \leq e-1$. The number of solutions of (36) prime to p is the number of sets y_1, \dots, y_n taken modulo $p-1$ which satisfy

$$(38) \quad \sum_{i=1}^n g^{a_i + e(q_i f + z_i)} \equiv d \pmod{p}.$$

Since $g^{ef} \equiv 1$, (38) reduces to (37) for each of the e^n sets q_1, \dots, q_n .

Let $n=2$, $a_1 = a_2 = 0$, $d=1$. Then (37) is

$$g^{ez_1} + g^{ez_2} \equiv 1, \quad 1 + g^{ev} \equiv g^{ew} \pmod{p}.$$

THEOREM 2. The number * of solutions prime to $p = ef + 1$ of

$$(39) \quad x^e + y^e \equiv 1 \pmod{p}$$

is $e^2(0, 0)$. The number of all solutions is $2e + e^2(0, 0)$.

THEOREM 3. For k and h chosen from $0, \dots, e-1$, the congruence

$$(40) \quad 1 + g^k x^e \equiv g^h y^e \pmod{p = ef + 1}$$

has exactly $e^2(k, h)$ solutions if $h \neq 0$ and

$$(41) \quad k \neq 0 \text{ if } f \text{ even}, \quad k \neq e/2 \text{ if } f \text{ odd};$$

$e + e^2(k, h)$ solutions if $h = 0$ and (41), or if $h \neq 0$ and

$$(42) \quad k = 0 \text{ if } f \text{ even}, \quad k = e/2 \text{ if } f \text{ odd};$$

$2e + e^2(k, h)$ solutions if $h = 0$ and (42).

By (5) and Theorem 1, (40) has exactly $e^2(k, h)$ solutions prime to p . The number of solutions with $x=0$ is e or 0 according as h is or is not divisible by e . Next, $y=0$ if and only if

$$-g^k \equiv g^{k+(p-1)/2}$$

is an e -th power, viz., $k + (p-1)/2$ divisible by e . When f is even, this is true only if k is divisible by e . When f is odd, $\frac{1}{2}(p-1) = \frac{1}{2}e + e(f-1)/2$, the condition is $k \equiv \frac{1}{2}e \pmod{e}$.

THEOREM 4. When f is even, the congruence

$$(43) \quad 1 + g^k x^e \equiv -g^h y^e \pmod{p = ef + 1}$$

* False result by G. Cornacchia, *Giornale di Matematico*, vol. 47 (1909), pp. 225, 235, 238, 241, etc.

has the same number of solutions as (40). When f is odd, the number of solutions is $N = e^2(k, h \pm \frac{1}{2}e)$ if $h \neq \frac{1}{2}e, k \neq \frac{1}{2}e$; $e + N$ if $h \neq \frac{1}{2}e, k = \frac{1}{2}e$, or $h = \frac{1}{2}e, k \neq \frac{1}{2}e$; $2e + N$ if $h = k = \frac{1}{2}e$.

When f is even, there exists an integer w belonging to the exponent $2e$, a divisor of $p - 1 = 2e \cdot f/2$, whence $w^e \equiv -1 \pmod{p}$.

When f is odd, $-g^h \equiv g^H \pmod{p}$, where $H = h - \frac{1}{2}e + e(f+1)/2$, and (43) is equivalent to

$$1 + g^k x^e \equiv g^{h-e/2} Y^e \pmod{p}.$$

Hence we apply Theorem 3 with h replaced by $h - \frac{1}{2}e$. The case $k = h = 0$ gives

THEOREM 5. If f is odd, the number of solutions of

$$(44) \quad 1 + x^e + y^e \equiv 0 \pmod{p = ef + 1}$$

is $e^2(0, \frac{1}{2}e)$. If f is even, it has the same number of solutions as (37).

THEOREM 6. If r, s, A are all prime to $p > 2$,

$$(45) \quad rx^2 + sy^2 = A \pmod{p}$$

has $p - N$ solutions,* where $N = +1$ or -1 according as $-rs$ is a quadratic residue or non-residue of p .

Since the theorem and the congruence are unaltered if we multiply r, s, A by the same integer prime to p , it suffices to prove the theorem for the case $A = -1$. Hence it suffices to prove that

$$(46) \quad 1 + rx^2 \equiv ty^2 \pmod{p}$$

has $p - N$ solutions, where $N = +1$ or -1 according as rt is a quadratic residue or non-residue of p .

Since a primitive root g is a quadratic non-residue of p , there are four cases: $r, t = 1$ or g . By Theorem 3 with $e = 2$, the number of solutions when f is odd is

$$\begin{aligned} &2 + 4(0, 0) \text{ if } r = t = 1, \quad 4(0, 1) \text{ if } r = 1, t = g; \\ &4 + 4(1, 0) \text{ if } r = g, t = 1; \quad 2 + 4(1, 1) \text{ if } r = t = g; \end{aligned}$$

* Jordan, *Traité des substitutions* (1870), pp. 156-161; *Comptes Rendus*, vol. 62 (1866), p. 687 (Lebesgue, *ibid.*, p. 868). The case of n variables is proved by induction on n .

but when f is even is $4 + 4(0, 0)$, $2 + 4(0, 1)$, $2 + 4(1, 0)$, $4(1, 1)$, in the respective cases. Applying (18) and (19), we obtain the statement below (46).

By Theorem 2 and (34),

$$(47) \quad x^3 + y^3 \equiv 1 \pmod{p = 3f + 1}$$

has exactly $p - 8 + L$ solutions prime to p . In case 2 is a cubic residue of p , $2y^3 \equiv 1 \pmod{p}$ has three roots and (47) has nine solutions prime to p with $x^3 \equiv y^3$. The solutions prime to p with $x^3 \not\equiv y^3$ fall into sets of $2 \times 3 \times 3$ (where those of a set have fixed values of x^3 and y^3 , also permuted). Hence $p - 8 + L \equiv 9 \pmod{18}$ and L is even. But if 2 is a cubic non-residue of p , $p - 8 + L \equiv 0 \pmod{18}$ and L is odd.

THEOREM 7. *Congruence (47) has $p - 2 + L$ solutions in all. 2 is a cubic residue of p if and only if L is even and $p = l^2 + 27m^2$ is then solvable.*

11. *Case $e = 4$.* Here $p = 4f + 1$. By (27) with $\beta^2 = -1$,

$$R(1, 1) = (00) - (01) + (02) - (03) - (20) + (21) - (22) + (23) \\ + 2\beta\{(10) - (11) + (12) - (13)\}.$$

Case $e = 4$, f even. For application to $e = 8$, we here write $[ij]$ for the usual (ij) . Then $[h, k] = [k, h]$ and

$$(48) \quad [13] = [23] = [12], [11] = [03], [22] = [02], [33] = [01].$$

Let L_1, L_2, L_3 denote the following equations, from (17):

$$(49) \quad [00] + [01] + [02] + [03] = f - 1, [01] + [03] + 2[12] = f, \\ [02] + [12] = \frac{1}{2}f.$$

$$L_1 - L_2 - L_3 : 3[12] = [00] + \frac{1}{2}f + 1,$$

$$L_1 - 2L_2 - 2L_3 : [00] - [01] - [02] - [03] - 6[12] = -2f - 1,$$

$$(50) \quad R(1, 1) = -x + 2\beta y, \quad x = 2f + 1 - 8[12], \quad y = [01] - [03].$$

$$(51) \quad p = x^2 + 4y^2, \quad x \equiv 1 \pmod{4}.$$

Here y is two-valued, depending on the choice of the primitive root g ; see below (93). We get

$$(52) \quad 16[00] = p - 11 - 6x, \quad 16[01] = h + 8y, \quad 16[02] = h, \\ 16[03] = h - 8y, \quad 16[12] = p + 1 - 2x, \quad h = p - 3 + 2x.$$

12. *Case $e = 4$, f odd.* By the correspondence (16), or direct,

$$(53) \quad \begin{aligned} 22 = 20 = 00, \quad 32 = 13 = 01, \quad 12 = 31 = 03, \\ 33 = 23 = 30 = 21 = 11 = 10, \end{aligned}$$

$$(54) \quad \begin{aligned} 00 + 01 + 02 + 03 = f, \quad 01 + 03 + 2(10) = f, \quad 00 + 10 = \frac{1}{2}(f-1), \\ R(1, 1) = -00 - 01 + 02 - 03 + 2(10) + 2\beta(03 - 01), \end{aligned}$$

Multiply (54) by $-1, 2, 2$ and add. Hence

$$(55) \quad R(11) = -x + 2\beta y, \quad x = 2f - 1 - 8(10), \quad y = (03) - (01).$$

Thus (51) holds. All the (ij) are determined; for example

$$(56) \quad (02) = 3(10) - \frac{1}{2}(f-1), \quad 16(02) = p + 1 - 6x.$$

13. *Case $e = 5$.* For a prime $p = 5f + 1$, f is even. For application to $e = 10$, write $[ij]$ for the usual (ij) . By (14),

$$(57) \quad 44 = 01, \quad 33 = 02, \quad 22 = 03, \quad 11 = 04, \quad 34 = 14 = 12, \quad 24 = 23 = 13,$$

and $[kh] = [hk]$. The twenty-five $[ij]$ reduce to $00, 01, 02, 03, 04, 12, 13$. Here (17) reduce to

$$(58) \quad \begin{aligned} 00 + 01 + 02 + 03 + 04 = f - 1, \quad 01 + 04 + 2[12] + 13 = f, \\ 02 + 03 + 12 + 2[13] = f. \end{aligned}$$

Let β be a primitive fifth root of unity, whence

$$(59) \quad \beta^4 + \beta^3 + \beta^2 + \beta + 1 = 0.$$

We eliminate the terms free of β from (27) and obtain

$$(60) \quad \begin{aligned} R(1, 1) &= a_1\beta + a_2\beta^2 + a_3\beta^3 + a_4\beta^4, \\ a_1 &= [02] - [00] + 2[01] - 2[12], \quad a_2 = [04] - [00] + 2[02] - 2[13], \\ a_3 &= [01] - [00] + 2[03] - 2[13], \quad a_4 = [03] - [00] + 2[04] - 2[12]. \end{aligned}$$

By (26) and (27),

$$\begin{aligned} p &= a_1^2 + a_2^2 + a_3^2 + a_4^2 + (\beta + \beta^4)B + (\beta^2 + \beta^3)C, \\ B &= a_1a_2 + a_2a_3 + a_3a_4, \quad C = a_1a_3 + a_2a_4 + a_1a_4. \end{aligned}$$

Replace $\beta + \beta^4$ by $-1 - \beta^2 - \beta^3$ and note that (59) is irreducible. Hence

$$(61) \quad p = a_1^2 + a_2^2 + a_3^2 + a_4^2 - B, \quad B = C.$$

Replacing B by $\frac{1}{2}(B + C)$, we see that

$$(62) \quad \begin{aligned} 16p &= x^2 + 5[a_1 - a_2 - a_3 + a_4]^2 + 10F^2 + 10G^2, \\ -x &= a_1 + a_2 + a_3 + a_4, \quad F = a_2 - a_3, \quad G = a_1 - a_4. \end{aligned}$$

By the values of the a_i , we get

$$(63) \quad \begin{aligned} x &= 25\{[12] + [13]\} - 10f - 4, \quad a_1 - a_2 - a_3 + a_4 = 5w, \\ w &= [13] - [12], \\ F &= 2u - v, \quad G = u + 2v, \quad u = [02] - [03], \quad v = [01] - [04]. \end{aligned}$$

Hence

$$(64) \quad 16p = x^2 + 50u^2 + 50v^2 + 125w^2,$$

and $x \equiv 1 \pmod{5}$. Using $B = C$, we find that

$$D = G^2 + 4FG - F^2 = (a_1 + a_4)^2 - (a_2 + a_3)^2 = -5xw.$$

Using $5u = 2F + G$, $5v = -F + 2G$, we find that

$$(65) \quad \begin{aligned} 25(u^2 + 4uv - v^2) &= 5D, \\ v^2 - 4uv - u^2 &= xw. \end{aligned}$$

By (58) and the value of x ,

$$(66) \quad \begin{aligned} [00] - 3[12] - 3[13] + f + 1 &= 0, \\ 25[00] &= p - 14 + 3x. \end{aligned}$$

Hence by Theorem 2,

$$(67) \quad X^5 + Y^5 \equiv 1 \pmod{p = 5f + 1} \text{ has } p - 4 + 3x \text{ solutions.}$$

By (62) and the definition (63) of w , we get

$$(68) \quad 4a_1, 4a_4 = 5w - x \pm 2G; \quad 4a_2, 4a_3 = -5w - x \pm 2F.$$

THEOREM 8. *There are exactly eight integral simultaneous solutions of (64) and (65). If (x, u, v, w) is one solution, also $(x, -u, -v, w)$ and $(x, \pm v, \mp u, -w)$ are solutions. The remaining four are derived from these four by changing all signs.*

I. Elementary proof. Since 5 is a quadratic residue of a prime $p = 5f + 1$, there are two roots of $s^2 \equiv 5 \pmod{p}$. We have

$$(64') \quad 50(u^2 + v^2) \equiv -x^2 - 125w^2 \pmod{p}.$$

In (65) transpose $4uv$, square and eliminate $u^4 + v^4$ by means of the square of (64'), and multiply by $s^2 \equiv 5$. We get

$$(69) \quad s(x^2 - 125w^2) \equiv 100(xw + 5uv) \pmod{p}.$$

From the products of (64') and (69) by 50 and 10,

$$(70) \quad 2500(u+v)^2 \equiv -1000xw - 50(x^2 + 125w^2) + 10s(x^2 - 125w^2).$$

Employ an integer r belonging to the exponent 5 modulo $p = 5f + 1$. Write $a = r^4 - r$, $b = r^3 - r^2$. Then $a^2 + b^2 \equiv -5$, $a^2 - b^2 \equiv s$, where $s = -r^4 + r^3 + r^2 - r$, $s^2 \equiv 5$, and $ab \equiv s$. Define $m = -2a - 4b$, $t = 4a - 2b$. Then

$$(71) \quad m^2 \equiv 10s - 50, \quad t^2 \equiv -10s - 50, \quad mt \equiv -20s.$$

Hence (70) is the square of either

$$(72) \quad 50(u+v) \equiv mx + 5stw \pmod{p}$$

or the like congruence in which the signs of u and v are both changed. This change is taken care of in the theorem.

Write $2K = t + m$, $2L = t - m$. Then (72) is the sum of

$$(73) \quad 50u \equiv Kx + 5sLw, \quad 50v \equiv -Lx + 5sKw \pmod{p}.$$

The product of (73) agrees with (69). The ambiguity in the determination of u and v is removed as follows. Replace r by r^2 and w by $-w$. Then K, L, s, u, v become $L, -K, -s, -v, u$ respectively. Hence (73) hold either for the given solution (x, u, v, w) or for the new solution $(x, -v, u, -w)$ of (64) and (65).

Let (x, u, v, w) and (x_1, u_1, v_1, w_1) be any integral solutions of (64), (65), (73). Evidently

$$xx_1 + 50uu_1 + 50vv_1 + 125ww_1 \equiv 0 \pmod{p}.$$

Denote the absolute value of the left member by A . By (64),

$$(16p)^2 = A^2 + 50(xu_1 - x_1u)^2 + 50(xv_1 - x_1v)^2 + 125(xw_1 - x_1w)^2 \\ + 2500(uv_1 - u_1v)^2 + 6250(uw_1 - u_1w)^2 + 6250(vw_1 - v_1w)^2.$$

Hence $A \leq 16p$, $6p^2 \equiv A^2 \pmod{25}$. By (64), $x \equiv w$, $x_1 \equiv w_1 \pmod{2}$. Hence $A = 2mp$,

$$4m^2 \equiv 6 \pmod{25}, \quad 2m = 5j \pm 1, \quad j \equiv \pm 3 \pmod{5}, \quad 2m \equiv \pm 16 \pmod{25}.$$

Hence $A = 16p$,

$$xu_1 - x_1u = 0, \quad \dots, \quad vw_1 - v_1w = 0.$$

Since (64) implies $x^2 \equiv x_1^2 \equiv 1 \pmod{5}$, $x \not\equiv 0$, $x_1 \not\equiv 0$,

$$u/x = u_1/x_1, \quad v/x = v_1/x_1, \quad w/x = w_1/x_1.$$

Hence (64) gives $x^2 = x_1^2$, whence $u^2 = u_1^2$, etc. This proves* Theorem 8.

Choose a definite one of the eight solutions of (64) and (65). Then the three linear equations (58) and the four linear equations whose left members are x, w, u, v uniquely determine $[0h]$, $h = 0, \dots, 4$, and $[12]$, $[13]$, and hence determine uniquely all 25 numbers $[i, j]$. This solves the cyclotomic problem for $e = 5$.

II. *Proof of Theorem 8 by algebraic numbers.* Let $p \equiv 1 \pmod{5}$. In the field F defined by an imaginary fifth root β of unity, the principal ideal (p) is the product† of four distinct prime ideals each of norm p . Since the class-number of F is 1, every ideal is a principal ideal. Hence

$$(p) = (p_1) \cdots (p_4), \quad p = Up_1 \cdots p_4,$$

where p_i is a polynomial in β^i with integral coefficients independent of i , and U is a unit. Write $f(\beta)$ for $p_1 p_2$. Then $f(\beta^4) = p_4 p_3$. The symmetric function $p_1 p_2 p_3 p_4$ is an integer I . Hence $p = UI$, $U = \pm 1$. Thus $\pm p = f(\beta)f(\beta^{-1})$. The lower sign is excluded by (62). Hence $U = 1$ and

$$(74) \quad p = f(\beta)f(\beta^4), \quad f(\beta) = a_1\beta + a_2\beta^2 + a_3\beta^3 + a_4\beta^4.$$

Similarly, $p_1 p_3 \cdot p_4 p_2$ furnishes a decomposition of p of type (74). But if $g(\beta) = p_1 p_4$, then $g(\beta^2) = p_2 p_3$ is not the product of $g(\beta^4) = g(\beta)$ by a unit, and we do not obtain a decomposition (74).

The replacement of β by β^3 yields $(p_1 p_3 p_4 p_2)$ and replaces (74) or $p_1 p_2 \cdot p_3 p_4$ by $p_3 p_1 \cdot p_4 p_2$ or $f(\beta^3)f(\beta^2)$, and gives rise to the substitution $S = (a_1 a_2 a_4 a_3)$. The replacement of β by β^4 of β^2 gives rise to the square $(a_1 a_4)(a_2 a_3)$ or cube $(a_1 a_3 a_4 a_2)$ of S .

Now S replaces x, u, v, w by $x, -v, u, -w$. Hence apart from the powers of S , the only decompositions of p into two conjugate factors are

$$p = Vf(\beta) \cdot V^{-1}f(\beta^4),$$

where V is a unit. Every unit of the field F is of the form

$$V = \pm \beta^k J^n, \quad J = \beta + \beta^4,$$

* In the much longer proof by G. Hull, *Transactions of the American Mathematical Society*, vol. 34 (1932), pp. 908-937, the sign of ξ in (87) should be changed. His y, z are our $u-v, -u-v$. Our u, v, w, x correspond to $C, D, A-B, \frac{1}{5}\{4p-16-25(A+B)\}$ of W. Burnside, *Proceedings of the London Mathematical Society*, (2), vol. 14 (1915), pp. 251-259.

† Kummer. Cf. Hilbert's Report, *Jahresbericht der Mathematischen Vereinigung*, Bd. 4 (1894-1895), pp. 328-329.

where k and n are integers. The condition for $V^{-1} = V(\beta^4)$ is $J^{-n} = J^n$. But

$$J^2 + J = 1, \quad 2J = -1 \pm 5^{1/2}, \quad |J| \neq 1.$$

Hence $n = 0$. If we change the sign of each factor in (74), we change the signs of each a_i and hence of x, u, v, w . We have now accounted by the eight solutions in Theorem 6.

It remains only to consider

$$p = \beta^k f(\beta) \cdot \beta^{4kf}(\beta^4).$$

In view of our examination of the effect of replacing β by β^k , it suffices to treat the case $k = 1$. For f in (74),

$$\beta f = A_1\beta + \cdots + A_4\beta^4, \quad A_1 = -a_4, \quad A_2 = a_1 - a_4, \quad A_3 = a_2 - a_4, \quad A_4 = a_3 - a_4.$$

Let V denote the function obtained from v by replacing a_i by A_i . By the analogue of $5v = -F + 2G$, we get

$$\begin{aligned} 5V &= -(A_2 - A_3) + 2(A_1 - A_4) = a_2 - a_1 - 2a_3, \\ 20V &= 2x + 2(3F - G) = 2x + 10u - 10v, \quad x \equiv 0 \pmod{5}, \end{aligned}$$

contrary to $x^2 \equiv 1$ by (64). Expressed otherwise, if a_i are integral solutions of (61) to which correspond integral solutions of (64) and (65), although the A_i evidently satisfy (61), the corresponding solutions X, \dots, W of (64) and (65) are not integers.

Example. $p = 11, a_1 = 0, a_2 = -1, a_3 = -2, a_4 = 2$. Then $x = w = 1, u = 0, v = -1; A_1 = A_2 = -2, A_3 = -3, A_4 = -4; X = 11, U = 4/5, V = 3/5, W = -1/5$.

14. *Subdivision of periods.* Let d be any divisor of e and write $E = e/d$. Then $(p-1)/E = df$. Replacing e, f by E, df in (1), we see that the E periods are

$$Y_k = \sum_{t=0}^{df-1} R^t e^{Et+k} \quad (k = 0, \dots, E-1).$$

The values $j, d+j, \dots, (f-1)d+j$ of t give the terms of η_{k+jE} in (1). Hence

$$(75) \quad Y_k = \sum_{j=0}^{d-1} \eta_{k+jE}.$$

Take $d = 2$. Then $e = 2E$ and

$$\begin{aligned} (76) \quad Y_k &= \eta_k + \eta_{k+E}, \\ Y_0 Y_k &= (\eta_0 + \eta_E)(\eta_k + \eta_{k+E}), \\ \eta_0 \eta_k &= f n_k + \sum_{h=0}^{E-1} (k, h) \eta_h + \sum_{H=0}^{E-1} (k, E+H) \eta_{E+H}, \end{aligned}$$

from which we get $\eta_0\eta_{k+E}$ by replacing k by $k + E$. Similarly

$$\eta_E\eta_{E+k} = fn_k + \sum_{h=0}^{E-1} (k, h)\eta_{E+h} + \sum_{H=0}^{E-1} (k, E+H)\eta_H,$$

by (7). Replacing m by k and k by $E - k$ in (7), we get

$$\eta_k\eta_E - fn_{E-k} = \sum_{h=E-k}^{2E-1-k} (E-k, h)\eta_{k+h} + \sum_{h=0}^{E-1-k} + \sum_{h=2E-k}^{2E-1}.$$

In the first sum, take $k + h = E + H$; we get

$$\sum_{H=0}^{E-1} (E-k, E+H-k)\eta_{E+H}.$$

In the second and third sums, take $k + h = H$. In the last case we may drop $2E$ from the subscripts of η . Combining, we get

$$\sum_{H=0}^{E-1} (E-k, H-k)\eta_H.$$

The total sum must be equal to (6) for Y periods:

$$Y_0Y_k = \sum_{h=0}^{E-1} (k, h)_E Y_h + 2fN_k, \quad Y_h = \eta_h + n_{h+E}.$$

By the coefficients of η_h ,

$$(77) \quad (k, h)_E = (k, h) + (k+E, h) + (k, E+h) + (E-k, h-k).$$

By way of check, we may verify that the coefficient of η_{h+E} in the total sum is also (77). By (14) and (15),

$$(78) \quad (00)_E = (00) + 3(0E), f \text{ even}; \quad (00)_E = 3(00) + (0E), f \text{ odd}.$$

In (22) for $e = 2E$, $m = 2M$, take $j = J + E$ in the terms with $j = E, \dots, 2E - 1$. By (76), we get.

$$F(\beta^{2M}) = \sum_{j=0}^{E-1} \beta^{2Mj} Y_j.$$

Now $B = \beta^2$ is a primitive E -th root of unity. Let $\phi(B^m)$ denote the function derived from $F(\beta^m)$ in (22) by replacing e by E , β by B , and η by Y . Hence $F(\beta^{2M}) = \phi(B^m)$. Applying (26) also for ϕ , we get

$$(79) \quad R(2r, 2s)_E = \{R(r, s)_E \text{ with } \beta \text{ replaced by } \beta^2\}.$$

15. *Jacobi's Theorem*.* If $g^m \equiv 2 \pmod{p}$, function (21) has the property

$$(80) \quad F(-1)F(\alpha^2) = \alpha^{2m}F(\alpha)F(-\alpha).$$

For i fixed, the coefficient of α^i in $F(\alpha)F(-\alpha)$ is

$$(81) \quad \sum_{j=0}^{p-2} (-1)^j R^c, \quad c \equiv g^{i-j} + g^j \pmod{p}.$$

Hence this is the coefficient of α^{2m+i} in the second member of (80). If i is odd, the sum (81) is zero since $j = J$ and $j = i - J$ give rise to the same value of c modulo p , while one of $J, i - J$ is even and the other is odd.

Henceforth, let i be even, $i = 2t$. Thus we seek the coefficient of α^{2m+2t} in $F(\alpha^2)$. It is obtained by replacing α by α^2 in the terms of $F(\alpha)$ in (21) having $k = m + t$ and $k = m + t + \frac{1}{2}(p-1)$. Hence the coefficient of α^{2m+2t} in $F(-1)F(\alpha^2)$ is

$$F(-1)(R^{g^{m+t}} + R^{-g^{m+t}}),$$

or, by use of $g^m \equiv 2$,

$$(82) \quad \sum_{k=0}^{p-2} (-1)^k R^{g^{k+2g^t}} + \sum_{k=0}^{p-2} (-1)^k R^{g^{k-2g^t}}.$$

First, let $J \not\equiv t \pmod{\frac{1}{2}(p-1)}$. Then J and $i - J$ are values of j incongruent modulo $p-1$, leading to the same c in (81); and the coefficient of R^c is $2(-1)^J$. The term R^c occurs in the first sum (82) for $g^k \equiv g^{-J}(g^t - g^J)^2$, and occurs in the second sum for $g^k \equiv g^{-J}(g^t + g^J)^2$. In each case g^k and g^J are both quadratic residues or both non-residues of p , whence $k \equiv J \pmod{2}$. Thus the coefficient of R^c in (82) is $2(-1)^J$.

Second, let $J \equiv t \pmod{\frac{1}{2}(p-1)}$. Then $g^J \equiv \pm g^t$, $c \equiv \pm 2g^t \pmod{p}$. Now only one of the two sums (82) yields a term R^c , the second when $g^k \equiv 4g^t$ and the upper sign holds, but the first when $g^k \equiv -4g^t$ and the lower sign holds. In both cases, $g^k \equiv 4g^J \pmod{p}$, $k \equiv J \pmod{2}$, whence $(-1)^J$ is the coefficient of R^c in both (81) and (82).

It remains to consider exponents c that do not occur in (81). Write z for g^j . Then $g^{2t}/z + z \equiv c \pmod{p}$ has no root z . Hence $c^2 - 4g^{2t}$ is a non-residue of p . Hence for one of the congruences

$$c - 2g^t \equiv g^k, \quad c + 2g^t \equiv g^k \pmod{p},$$

the solution g^k is a residue and for the other a non-residue. Thus x^c occurs in one of the sums (82) with the coefficient $+1$ and in the other with -1 .

* Stated without proof in *Journal für Mathematik*, Bd. 30 (1846), p. 167. The present proof was recently obtained by H. H. Mitchell.

Since we have found that (81) and (82) have the same coefficients of R^c for every c , we have proved (80).

16. *The reduced $R(m, n)$.* By (25) and (26)

$$(83) \quad R(n, m) = R(m, n) = (-1)^{nf} R(-m - n, n),$$

when no one of $m, n, m + n$ is divisible by e .

When β is replaced by a new primitive e -th root β^j of unity (j prime to e), $R(m, n)$ becomes $R(jm, jn)$. The latter is called a conjugate of the former. The relation obtained from (28) by this replacement evidently yields the same decomposition of p into integers that (28) itself yields.

When we retain only one of a set of conjugate R 's and discard duplicates by (83), we obtain a set of reduced R 's.

Examples of complete sets of reduced R 's:

$$e = 6 : R(1, 1); R(1, 2), R(2, 2).$$

$$e = 8 : R(1, 1), R(1, 3), R(1, 5), R(2, 2).$$

17. **THEOREM 9.** *When $e = 6$, the 36 cyclotomic constants (k, h) depend solely upon the decomposition $A^2 + 3B^2$ of the prime $p = 6f + 1$.*

By (83), $R(11) = (-1)^f R(14)$. Employ the values of $R(14)$ and $R(12)$ from (26) and apply (80) for $\alpha = \beta^2$; we get the first of

$$(84) \quad R(11) = (-1)^f \beta^{4m} R(12), \quad R(22) = \beta^{2m} R(12),$$

the second of which follows from (80) for $\alpha = \beta$. Here

$$\beta^3 = -1, \quad 2\beta = 1 + (-3)^{\frac{1}{2}}, \quad 2\beta^2 = -1 + (-3)^{\frac{1}{2}}.$$

By (79) and (32), we get the first of

$$(85) \quad \begin{aligned} 2R(22) &= L + 3M(-3)^{\frac{1}{2}}, & R(12) &= -A + B(-3)^{\frac{1}{2}}, \\ 2R(11) &= E + F(-3)^{\frac{1}{2}}. \end{aligned}$$

18. *Case $e = 6, f$ even.* Since our results will be needed for $e = 12$, we shall here write $[ij]$ for the usual (ij) . Then

$$(86) \quad \begin{aligned} [kh] &= [hk], & [01] &= [55], & 02 &= 44, & 03 &= 33, & 04 &= 22, \\ 05 &= 11, & 12 &= 15 = 45, & 13 &= 25 = 34, & 14 &= 23 = 35. \end{aligned}$$

We retain the first one in each equation and $[00], [24]$. Then (17) reduce to

$$(87) \quad \begin{aligned} 00 + 01 + 02 + 03 + 04 + 05 &= f - 1, \\ 01 + 05 + 2[12] + 13 + 14 &= f, \\ 02 + 04 + 12 + 13 + 14 + 24 &= f, & 03 + 13 + 14 &= \frac{1}{2}f. \end{aligned}$$

Multiply these by $-1, 1, 1, -2$, and add; we get

$$(88) \quad [24] - [00] - 3[03] + 3[12] = 1.$$

$$(89) \quad \begin{aligned} A &= 2[24] - 2[00] - 1, & B &= [01] - [05] - [13] + [14], \\ E &= 2[00] - 6[03] - 3[12] + 7[24], \\ F &= [01] + 2[13] - 3[04] + 3[02] - [05] - 2[14]. \end{aligned}$$

Thus $A \equiv 1 \pmod{6}$. By (32), (78); (31₂), (77); and (33)

$$(90) \quad \begin{aligned} 9[00] + 27[03] &= p - 8 + L, & 4p &= L^2 + 27M^2, & L &\equiv 1 \pmod{3}, \\ M &= [01] + 2[13] + [04] - [02] - 2[14] - [05]. \end{aligned}$$

I. Let 2 be a cubic residue of p . Then $\beta^{2m} = 1$. By (84), (85), $L = E = -2A$, $F = 2B = 3M$. Then (87)-(90) give

$$(91) \quad \begin{aligned} 36[00] &= p - 17 - 20A, & 36[03] &= \pi, & 36[12] &= p + 1 - 2A, \\ 36[01] &= \pi + 18B, & 36[05] &= \pi - 18B, & 36[02] &= \pi + 6B, \\ 36[04] &= \pi - 6B, & [13] &= [14] = [24] = [12], & \pi &= p - 5 + 4A. \end{aligned}$$

II. In $g^m \equiv 2 \pmod{p}$, let $m \equiv 2$ or $5 \pmod{6}$. Then

$$(92) \quad \begin{aligned} E &= A - 3B, & F &= -A - B, & L &= A + 3B, & 3M &= A - B, \\ 36[00] &= p - 17 - 8A - 6B, & 36[03] &= \pi + 6B, & 36[24] &= p + 1 + 10A - 6B, \\ 36[12] &= 36[14] = p + 1 - 2A + 6B, & 36[05] &= \pi - 12B, & [04] &= [01] = [03], \\ 36[13] &= p + 1 - 2A - 12B, & 36[02] &= p - 5 - 8A, & \pi &= p - 5 + 4A. \end{aligned}$$

III. Let $m \equiv 1$ or $4 \pmod{6}$. Then $E = A + 3B$, $F = A - B$, $L = A - 3B$, $3M = -A - B$, and

$$(93) \quad \begin{aligned} 36[00] &= p - 17 - 8A + 6B, & 36[03] &= 36[02] = 36[05] = \pi - 6B, \\ 36[12] &= 36[13] = p + 1 - 2A - 6B, & 36[24] &= p + 1 + 10A + 6B, \\ 36[14] &= p + 1 - 2A + 12B, & 36[01] &= \pi + 12B, & 36[04] &= p - 5 - 8A. \end{aligned}$$

We may deduce case III from II as follows.

For any e, f , replace g by a new primitive root g^r of p , where r is prime to $p - 1$. Then η_k in (1) becomes η_{rk} since rt ranges with t over a complete set of residues modulo f . By (6), (k, h) becomes (rk, rh) .

Let $r'r \equiv 1 \pmod{e}$. Since $rj = J$ ranges with j over a complete set of residues modulo e , $F(\beta^m)$ in (22) becomes

$$\sum_{j=0}^{e-1} \beta^{mr'J} \eta_J = F(\beta^{mr'}).$$

By (26), $R(m, n)$ becomes $R(mr', nr')$.

THEOREM 10. When g is replaced by a new primitive root g^r , $R(m, n)$ becomes $R(mr', nr')$, where $r'r \equiv 1 \pmod{e}$. The effect on any F or R is to replace β by β^r .

For our case $e = 6$, f even, take $r \equiv 5 \pmod{6}$. The replacement of β by β^{-1} is equivalent to changing the sign of $(-3)^{\frac{1}{2}}$. Hence by (85), $A, E, L, 00, 03, 12$ and 24 are unaltered, while B, F, M are changed in sign. But 01 and $05, 02$ and $04, 13$ and 14 are interchanged. Then (92) become (93) and conversely.

19. Case $e = 6, f$ odd. We retain 03 and the first one in each equation

$$\begin{aligned}
 (94) \quad & 00 = 30 = 33, \quad 01 = 25 = 43, \quad 02 = 14 = 53, \quad 21 = 45, \\
 & 04 = 13 = 52, \quad 05 = 23 = 41, \quad 10 = 22 = 31 = 34 = 40 = 55, \\
 & 11 = 20 = 32 = 35 = 44 = 50, \quad 15 = 12 = 24 = 42 = 51 = 54; \\
 (95) \quad & 00 + 01 + 02 + 03 + 04 + 05 = f, \quad 02 + 04 + 10 + 11 + 2(15) = f, \\
 & 01 + 05 + 10 + 11 + 15 + 21 = f, \quad 00 + 10 + 11 = \frac{1}{2}(f-1); \\
 & (21) - (03) - 3(00) + 3(15) = 1,
 \end{aligned}$$

$$\begin{aligned}
 L &= 27(00) + 9(03) - p + 8, \quad M = 01 + 04 + 2(10) - 02 - 05 - 2(11), \\
 (96) \quad & A = 2(03) - 2(21) \equiv 4 \pmod{6}, \quad B = 10 - 11 + 02 - 04, \\
 & E = 2(03) - 6(00) - 3(12) + 7(21) - 2, \\
 & F = 04 - 3(01) + 2(10) - 02 + 3(05) - 2(11).
 \end{aligned}$$

While L and M are the same functions of A, B as in I-III, E and F are the negatives of their former functions of A, B .

I. $2 = \text{cubic residue of } p$. For $t = p + 1 - 2A$,

$$(97) \quad 36(00) = p - 11 - 8A, \quad 36(03) = t + 18A, \quad 36(15) = 36(21) = t.$$

II. In $g^m \equiv 2 \pmod{p}$, let $m \equiv 2$ or $5 \pmod{6}$, $q = p + 1 + A + 3B$. Then

$$\begin{aligned}
 (98) \quad & 36(00) = p - 11 - 2A, \quad 36(03) = q + 9A + 9B. \\
 & 36(15) = q + 3A - 3B, \quad 36(21) = q - 9A + 9B.
 \end{aligned}$$

III. If $m \equiv 1$ or $4 \pmod{6}$, change the sign of B in II.

20. THEOREM 11. When $e = 8$, the 64 cyclotomic constants (k, h) depend solely upon the decompositions $p = x^2 + 4y^2$ and $p = a^2 + 2b^2$, $x \equiv a \equiv 1 \pmod{4}$.

Here $p = 8f + 1$, $\beta^4 = -1$, $(\beta + \beta^3)^2 = -2$. By (26) and (80) for $\alpha = \beta^3$, we get $R(16) = \beta^{6m}R(13)$. Next, employ (80) for $\alpha = \beta$ and divide by $F(\beta^6)$. Hence $R(24) = \beta^{2m}R(15)$. Applying (83), we get

$$(99) \quad R(22) = \beta^{2m}R(15), \quad R(11) = (-1)^f \beta^{6m}R(13), \quad g^m \equiv 2 \pmod{p}.$$

21. Case $e = 8$, f even. By (14), $(hk) = (kh)$ and

$$\begin{aligned} 11 &= 07, \quad 17 = 12, \quad 22 = 06, \quad 23 = 16, \quad 26 = 24, \quad 27 = 13, \quad 33 = 05, \\ 34 &= 15, \quad 35 = 25, \quad 36 = 25, \quad 37 = 14, \quad 44 = 04, \quad 45 = 14, \quad 46 = 24, \\ 47 &= 15, \quad 55 = 03, \quad 56 = 13, \quad 57 = 16, \quad 66 = 02, \quad 67 = 12, \quad 77 = 01. \end{aligned}$$

Hence each of the sixty-four (ij) is equal to one of the fifteen:

$$(100) \quad (0h), h = 0, \dots, 7; \quad (1h), h = 2, \dots, 6; \quad (24), (25).$$

Write $[ij]$ for $(ij)_4$. Their values are given by (52). Then by (77),

$$\begin{aligned} (101) \quad (00) &= [00] - 3(04), \quad (01) = [01] - (05) - 2(14), \\ (02) &= [02] - (06) - 2(24), \quad (03) = [03] - (07) - 2(15), \\ (12) &= [12] - (13) - (16) - (25). \end{aligned}$$

Eliminating the left members from (17), we get

$$\begin{aligned} (102) \quad [00] + [01] + [02] + [03] &= 2f - 1, \quad [02] + [12] = f, \\ [01] + 2[12] + [03] &= 2f, \\ (07) &= [03] + (05) + (13) + (14) - (15) + (16) + 2(25) - f, \\ (04) + (14) + (15) + (24) &= \frac{1}{2}f, \end{aligned}$$

the first three of which are (49) with f replaced by $2f$. By (79), (50), (51),

$$(103) \quad R(22) = -x + 2\beta^2y, \quad p = x^2 + 4y^2, \quad x \equiv 1 \pmod{4}.$$

From $R(mn)$ in (27) we eliminate the left members of (101) and (07) by (102), and get

$$\begin{aligned} R(13) &= -a + b(\beta + \beta^3), \quad p = a^2 + 2b^2, \\ -a &= [00] - [01] + [02] - [03] - 4(04) + 4(14) + 4(15) - 4(24), \\ b &= [01] - [03] - 4(05) - 4(13) - 4(14) - 4(25) + 2f. \end{aligned}$$

$$\begin{aligned} R(15) &= A + \beta^2B, \quad p = A^2 + B^2, \\ A &= [00] - [02] - 2[12] + 4\{- (04) + (13) + (16) + (24)\}, \\ B &= [01] + 2[02] - [03] + 4\{- (06) - (14) + (15) - (24)\}. \end{aligned}$$

$$R(11) = A_0 + 2A_1\beta + A_2\beta^2 + 2A_3\beta^3,$$

$$A_0 = [00] - [02] + 2[12] + 4\{-(04) - (13) - (16) + (24)\},$$

$$A_1 = [01] - [12] + 2\{-(05) - (14) + (13) + (25)\},$$

$$A_2 = -[01] + 2[02] + [03] + 4\{-(06) + (14) - (15) - (24)\},$$

$$A_3 = -[03] + [12] - 2(05) - 2(13) - 2(14) - 4(16) - 6(25) + 2f.$$

All the (ij) are uniquely determined by our linear equations and (52). By the first and last of (102), we get

$$(104) \quad 4[00] - 16(04) = A + A_0 - a - 1.$$

This with the first of (101) yield (00) and (04). Other simple relations are

$$(105) \quad A_1 - A_3 = 4\{(25) - (12)\}, \quad A_2 + B = 4\{(02) - (06)\}.$$

Since 2 is a quadratic residue of $p = 8f + 1$, there are two cases.

I. Let 2 be a biquadratic residue of p , whence m is a multiple of 4 and $\beta^{2m} = +1$. Then (99) gives

$$(106) \quad A = -x, \quad B = 2y, \quad A_0 = -a, \quad 2A_1 = 2A_3 = b, \quad A_2 = 0.$$

Then (52), (104) and (101₁) give

$$(107) \quad 64(00) = p - 23 - 18x - 24a, \quad 64(04) = p - 7 - 2x + 8a.$$

II. Let 2 be a biquadratic non-residue of p , whence m is the double of an odd integer, and $\beta^{2m} = -1$. Then by (99),

$$(108) \quad A = x, \quad B = -2y, \quad A_0 = a, \quad 2A_1 = 2A_3 = -b, \quad A_2 = 0.$$

$$(109) \quad 64(00) = p - 23 + 6x, \quad 64(04) = p - 7 - 10x.$$

Examples. If $p = 17$, $(02) = (15) = (16) = 1$; the others of (100) are zero.

p	00	01	02	03	04	05	06	07
97	2	2	0	2	0	2	1	2
113	0	0	0	2	3	2	2	4
257	9	2	2	6	2	0	6	4

p	12	13	14	15	16	24	25
97	1	0	2	1	3	3	1
113	1	3	1	1	3	2	1
257	5	3	5	5	3	4	5

22. Case $e = 8, f$ odd. Here

$$\begin{aligned} 14 = 05, 13 = 16, 15 = 03, 22 = 20, 23 = 17, 24 = 06, 25 = 16, 26 = 02, \\ 27 = 12, 30 = 11, 31 = 32 = 21, 33 = 10, 34 = 07, 35 = 17, 36 = 12, \\ 37 = 01, 40 = 00, 41 = 10, 42 = 20, 43 = 11, 44 = 00, 45 = 10, 46 = 20, \\ 47 = 11, 50 = 10, 51 = 07, 52 = 17, 53 = 12, 54 = 01, 55 = 11, 56 = 21, \\ 57 = 21, 60 = 20, 61 = 17, 62 = 06, 63 = 16, 64 = 02, 65 = 12, 66 = 20, \\ 67 = 21, 70 = 11, 71 = 12, 72 = 16, 73 = 05, 74 = 03, 75 = 16, 76 = 17, \\ 77 = 10. \end{aligned}$$

Write $[ij]$ for $(ij)_4, j$ taken modulo 4. By (77),

$$\begin{aligned} (110) \quad (04) &= [00] - 3(00), \quad (05) = [01] - (01) - 2(10), \\ (06) &= [02] - (02) - 2(20), \quad (07) = [03] - (03) - 2(11), \\ (16) &= [12] - (12) - (17) - (21). \end{aligned}$$

Eliminate the left members from (17); we get the first three in (102) and

$$\begin{aligned} (111) \quad 03 &= [03] + 01 + 10 - 11 + 12 + 17 + 2(21) - f, \\ 00 + 10 + 11 + 20 &= \frac{1}{2}(f - 1). \end{aligned}$$

The formulas for a, \dots, A_3 are derived from those for f even by replacing (k, h) by $(k, h + 4)$, from (16), where entries ≥ 8 are to be reduced modulo 8. The $[ij]$ are unaltered. We change the sign of a , so that the new a shall be $\equiv 1 \pmod{4}$ for f odd or even. As before,

$$(112) \quad 4[00] - 16(00) = A + A_0 + a + 1.$$

I. Let 2 be a biquadratic residue of p . Then

$$(113) \quad A = -x, \quad B = 2y, \quad A_0 = -a, \quad 2A_1 = 2A_3 = -b, \quad A_2 = 0,$$

$$(114) \quad 64(00) = p - 15 - 2x; \quad 64(04) = p + 1 - 18x.$$

II. Let 2 be a biquadratic non-residue of p . Then the second members of (113) are to be changed in sign. Thus

$$(115) \quad 64(00) = p - 15 - 10x - 8a, \quad 64(04) = p + 1 + 6x + 24a.$$

23. Case $e = 10$. Then $\beta^5 = -1$. By (80) with $\alpha = \beta^2$,

$$(116) \quad F(\beta^5)F(\beta^4) = \beta^{4m}F(\beta^2)F(\beta^7).$$

Divide by $F(\beta^9)$ and apply (26). Thus $R(45) = \beta^{4m}R(27)$. By (83),

$$R(45) = R(14), \quad R(27) = R(12), \quad R(14) = \beta^{4m}R(12).$$

By (80) with $\alpha = \beta^4$, $R(18) = \beta^{8m}R(14)$. Thus

$$R(18) = \beta^{2m}R(27), \quad F(\beta)F(\beta^8) = \beta^{2m}F(\beta^2)F(\beta^7).$$

Hence by (116), $\beta^{2m}F(\beta)F(\beta^8) = F(\beta^5)F(\beta^4)$. Multiplication by $F(\beta^4)/F(\beta^5)F(\beta^8)$ yields $\beta^{2m}R(14) = R(44)$. By (83), $R(11) = (-1)^f R(18)$. Hence

$$(117) \quad R(14) = \beta^{8m}R(44), \quad R(12) = \beta^{4m}R(44), \quad R(11) = (-1)^f \beta^{6m}R(44).$$

These four R 's are the only reduced ones. By (79), (62) and the remark below (83),

$$(118) \quad R(44) = -a_4\beta + a_3\beta^2 - a_2\beta^3 + a_1\beta^4.$$

We shall employ the notations

$$(119) \quad \begin{aligned} R(11) &= b_1\beta + b_2\beta^2 + b_3\beta^3 + b_4\beta^4, \\ R(12) &= d_1\beta + \dots, \quad R(14) = c_1\beta + \dots + c_4\beta^4. \end{aligned}$$

Since p in (62) is the product of (60) by its conjugate, by changing β to $-\beta$, we obtain the present analogue of (62) by changing the signs of a_1 and a_3 . Just as $(00)_5$ is determined by (66) from $-x = \sum a_i$, we shall find here that (00) and (05) are determined by

$$(120) \quad \begin{aligned} \rho(11) &= -b_1 + b_2 - b_3 + b_4, \quad \rho(12) = -d_1 + d_2 - d_3 + d_4, \\ \rho(14) &= -c_1 + c_2 - c_3 + c_4. \end{aligned}$$

I. $m \equiv 0 \pmod{5}$. By (117)-(120)

$$\rho(11) = (-1)^f(a_1 + a_2 + a_3 + a_4), \quad \rho(12) = \rho(14) = a_1 + a_2 + a_3 + a_4.$$

II. $2m \equiv 2 \pmod{10}$. Eliminate constant terms by

$$\beta^4 - \beta^3 + \beta^2 - \beta + 1 = 0.$$

Then

$$\begin{aligned} \rho(11) &= (-1)^f(a_2 + a_3 + a_4 - 4a_1), \quad \rho(12) = a_1 + a_2 + a_3 - 4a_4, \\ \rho(14) &= a_1 + a_2 + a_4 - 4a_3. \end{aligned}$$

III. $2m \equiv 4 \pmod{10}$. $\rho(11) = (-1)^f(a_1 + a_3 + a_4 - 4a_2)$,

$$\rho(12) = a_1 + a_2 + a_4 - 4a_3, \quad \rho(14) = a_2 + a_3 + a_4 - 4a_1.$$

IV. $2m \equiv 6 \pmod{10}$. $\rho(11) = (-1)^f(a_1 + a_2 + a_4 - 4a_3)$,

$$\rho(12) = a_1 + a_3 + a_4 - 4a_2, \quad \rho(14) = a_1 + a_2 + a_3 - 4a_4.$$

$$\text{V. } 2m \equiv 8 \pmod{10}. \quad \rho(11) = (-1)^f(a_1 + a_2 + a_3 - 4a_4),$$

$$\rho(12) = a_2 + a_3 + a_4 - 4a_1, \quad \rho(14) = a_1 + a_3 + a_4 - 4a_2.$$

24. Case $e = 10$, f even. We have $(h, k) = (k, h)$ and

$$\begin{aligned} 11 &= 09, & 19 &= 12, & 22 &= 08, & 23 &= 18, & 28 &= 24, & 29 &= 13, & 33 &= 07, & 34 &= 17, \\ 35 &= 27, & 37 &= 36, & 38 &= 25, & 39 &= 14, & 44 &= 06, & 45 &= 16, & 46 &= 26, & 47 &= 36, \\ 48 &= 26, & 49 &= 15, & 55 &= 05, & 56 &= 15, & 57 &= 25, & 58 &= 27, & 59 &= 16, & 66 &= 04, \\ 67 &= 14, & 68 &= 24, & 69 &= 17, & 77 &= 03, & 78 &= 13, & 79 &= 18, & 88 &= 02, & 89 &= 12, \\ 99 &= 01. \end{aligned}$$

Denote $(kh)_s$ by $[k, h]$ as in § 13. By (77),

$$\begin{aligned} (121) \quad (00) &= [00] - 3(05), & (01) &= [01] - (06) - 2(15), \\ (02) &= [02] - (07) - 2(25), & (03) &= [03] - (08) - 2(27), \\ (04) &= [04] - (09) - 2(16), & (12) &= [12] - (14) - (17) - (26), \\ (13) &= [13] - (18) - (24) - (36). \end{aligned}$$

The linear relations (17) reduce to (58) with f replaced by $2f$, and

$$\begin{aligned} (122) \quad (08) &= [03] + [13] + (07) + (14) + (17) - (24) + (25) \\ &\quad - (27) + (36) - f, \\ (09) &= [04] + (06) + (14) + (15) - (16) + (17) + (24) \\ &\quad + 2(26) + (36) - f, \\ (123) \quad (05) &+ (15) + (16) + (25) + (27) = \frac{1}{2}f. \end{aligned}$$

In (119) we have

$$\begin{aligned} b_1 &= 00 - 02 - 05 - 07 + 2\{01 - 06 + 12 - 14 - 17 - 2(18) + 25 \\ &\quad + 26 + 2(36)\}, \\ b_2 &= 04 - 00 + 05 + 09 + 2\{02 - 07 - 2(12) + 13 + 2(14) - 16 + 18 \\ &\quad - 24 - 36\}, \\ b_3 &= 00 - 01 - 05 - 06 + 2\{03 - 08 + 2(12) - 13 + 15 - 2(17) - 18 \\ &\quad + 24 + 36\}, \\ b_4 &= 03 - 00 + 05 + 08 + 2\{04 - 09 - 12 + 2(13) + 14 + 17 - 26 \\ &\quad - 27 - 2(36)\}. \end{aligned}$$

Eliminate the left members of (121) and (122); subtract (17) for $k=0$; and add the double of (123); we get

$$\begin{aligned} (124) \quad \rho(11) &= 20(05) - 1 - 5[00] + 2\{-[01] + [02] + [03] - [04] \\ &\quad - 6[12] + 6[13]\} + 20\{(14) + (17) - (24) - (36)\}. \end{aligned}$$

Write $z_j = \Sigma (-1)^h(jh)$, $h = 0, \dots, 9$. Then in R_{14} ,

$$\begin{aligned} c_1 &= z_0 + z_1 - z_4, & c_2 &= -z_0 + z_2 + z_3, & c_3 &= z_0 - z_2 + z_3, \\ c_4 &= -z_0 + z_1 + z_4, \\ (125) \quad \rho(14) &= -4z_0 + 2z_2 + 2z_4 = -4[00] + 4[01] - 2[02] - 2[03] \\ &\quad + 4[04] - 2[12] - 8[13] - 2f + 20\{(05) + (24) + (26)\}, \end{aligned}$$

after adding the product of (123) by 4. Next,

$$R(12) = t_0 + t_1\beta + \dots + t_4\beta^4,$$

where

$$\begin{aligned} t_0 &= [00] - 2[12] - 4(05) + 2(14) + 2(17) + 2(24) + 4(26) - 2(36), \\ t_1 &= [01] - 2(06) - 2(08) - 2(15) + 2(17) - 2(26) + 2(27), \\ t_2 &= [02] + 2(06) - 2(07) - 2(15) - 2(18) + 2(24) - 2(25), \\ t_3 &= [03] - 2[04] + 2[13] - 2(08) + 2(09) + 6(16) - 2(18) - 4(24) \\ &\quad - 2(27) - 2(36), \\ t_4 &= 2[02] + [04] - 2(07) - 2(09) - 2(14) - 2(16) - 6(25) + 2(26), \\ d_1 &= t_1 + t_0, \quad d_2 = t_2 - t_0, \quad d_3 = t_3 + t_0, \quad d_4 = t_4 - t_0. \end{aligned}$$

$$\begin{aligned} (126) \quad \rho(12) &= -4t_0 - t_1 + t_2 - t_3 + t_4 \\ &= -4[00] - [01] + 3[02] + 3[03] - [04] + 8[12] + 2[13] \\ &\quad - 2f + 20(05) - 20(26) - 10\{(14) + (17) + (24) - (36)\}, \end{aligned}$$

after adding the product of (123) by 4. By (58) we find that

$$(127) \quad \rho(11) + 2\rho(12) + 2\rho(14) = 100(05) - 25[00] - 5.$$

This with (121₁), (66) and (68) determine (05) and (00).

- I. $m \equiv 0 \pmod{5}$. $100(05) = p - 9 - 2x$, $100(00) = p - 29 + 18x$.
- II. $2m \equiv 2 \pmod{10}$. $400(05) = 4p - 36 + 17x + 50u - 25w$,
 $400(00) = 4p - 116 - 3x - 150u + 75w$.
- III. $2m \equiv 4 \pmod{10}$. $400(05) = 4p - 36 + 17x - 50v + 25w$,
 $400(00) = 4p - 116 - 3x + 150v - 75w$.
- IV. $2m \equiv 6 \pmod{10}$. Change the sign of v in III.
- V. $2m \equiv 8 \pmod{10}$. Change the sign of u in II.

25. *Case $e = 10$, f odd.* By means of the correspondence (16), which leaves the $[ij]$ unaltered, we may deduce from the results for f even the equalities between the (ij) , and the analogues to (121), (122), b_i , c_i , t_i . But (123) is here replaced by

$$(128) \quad (00) + (10) + (11) + (20) + (22) = \frac{1}{2}(f-1).$$

The present $\rho(11)$ is therefore derived from (124) by replacing -1 by $+1$. But for $k=2$ or 4 , the present $\rho(1k)$ is obtained by subtracting 2 from the former $\rho(1k)$. Hence

$$(129) \quad \rho(11) - 2\rho(12) - 2\rho(14) = 100(00) - 25[00] + 5.$$

- I. $m \equiv 0 \pmod{5}$. $100(00) = p - 19 + 8x$, $100(05) = p + 1 - 12x$.
 II. $2m \equiv 2 \pmod{10}$. $400(00) = 4p - 76 + 7x - 50u + 25w$,
 $400(05) = 4p + 4 + 27x + 150u - 75w$.
 III. $2m \equiv 4 \pmod{10}$. $400(00) = 4p - 76 + 7x + 50v - 25w$,
 $400(05) = 4p + 4 + 27x - 150v + 75w$.
 IV. $2m \equiv 6 \pmod{10}$. Change the sign of v in III.
 V. $2m \equiv 8 \pmod{10}$. Change the sign of u in II.

26. THEOREM 12. When $e=12$, the 144 cyclotomic constants (k, h) depend solely upon the decompositions $p = x^2 + 4y^2$ and $p = A^2 + 3B^2$ of the prime $p = 12f + 1$, where $x \equiv 1 \pmod{4}$, $A \equiv 1 \pmod{6}$.

As the reduced R 's we may take $R(1k)$, $k=1, 2, 3, 5, 7$, $R(22)$, $R(24)$, $R(33)$, $R(44)$. By (80) with $\alpha = \beta, \beta^2$ or β^5 ,

$$R(26) = \beta^{2m}R(17), \quad R(46) = \beta^{4m}R(28), \quad R(1, 10) = \beta^{10m}R(15).$$

By (83), $R(26) = R(46)$, $R(28) = R(22)$, $R(11) = (-1)^f R(1, 10)$.

Hence

$$(130) \quad R(17) = \beta^{2m}R(22), \quad R(11) = (-1)^f \beta^{10m}R(15).$$

Jacobi (*loc. cit.*) stated a formula involving an imaginary cube root γ of unity. Thus $\gamma = \beta^4$ or β^8 . For either, his formula becomes

$$(131) \quad F(\alpha)F(\beta^4\alpha)F(\beta^8\alpha) = \alpha^{-3m'}pF(\alpha^3), \quad g^{m'} \equiv 3 \pmod{p}.$$

We employ this only for $\alpha = \beta^9$ or β^5 , and eliminate p by (25) with $n=9$ or 5 . By (83), $R(19) = (-1)^f R(12)$. Hence

$$(132) \quad R(15) = (-1)^f k R(33), \quad R(12) = k R(37), \quad k = \beta^{-3m'}.$$

Since $p = 12f + 1$, 3 is a quadratic residue of p , while 2 is a quadratic residue (m even) or a non-residue (m odd), according as f is even or odd, whence

$$(133) \quad m' \text{ is even, } k^2 = 1; \quad \beta^{6m} = (-1)^f.$$

In $p = 6f' + 1$, f' is even. By (84) with $f = f'$, (79) and (130), (83),

$$(134) \quad R(22) = R(44), \quad R(17) = (-1)^f R(44), \quad R(14) = R(44).$$

By (26), the last gives $R(18) = R(45)$. Let $R(13) = cR(15)$. Then, by (26), $R(36) = cR(45)$. By (83), $R(36) = (-1)^f R(33)$, $R(18) = (-1)^f R(13)$. Hence $R(33) = cR(13) = c^2 R(15)$. Then (132) gives $(-1)^f k c^2 = 1$. By (133),

$$(135) \quad \text{either } *k = (-1)^f, c = \mp 1; \text{ or } k = -(-1)^f, c = \pm \beta^3; \\ R(13) = cR(15).$$

Then by (130₂), $R(13) = dR(11)$, $d = (-1)^f c \beta^{2m}$. By (26), $R(23) = dR(14)$. By (83), $R(37) = dR(17)$. By (130), (132),

$$(136) \quad R(12) = (-1)^f c k \beta^{4m} R(22).$$

By (79) and (85),

$$(137) \quad 2R(22) = E + F(2\beta^2 - 1), \quad 2R(44) = L + 3M(2\beta^2 - 1).$$

By (75) with $d = 3$, $E = 4$, $e = 12$, we find that

$$(138) \quad (0j)_4 = (0j) + (4j) + (8j) + (8, 8 + j) + (0, 8 + j) \\ + (4, 8 + j) + (4, 4 + j) + (8, 4 + j) + (0, 4 + j).$$

27. Case $e = 12$, f odd. By (15),

$$\begin{aligned} 16 = 07, 17 = 05, 18 = 15, 25 = 19, 26 = 08, 27 = 15, 28 = 04, 29 = 14; \\ 2, 10 = 24, 33 = 30, 34 = W, 35 = Z, 36 = 09, 37 = 19, 38 = 14, 39 = 03; \\ 3, 10 = 13; 3, 11 = 23, 40 = 22, 41 = 32, 43 = 31, 44 = 20, 45 = V, 46 = X, \\ 47 = Z, 48 = 24, 49 = 13; 4, 10 = 02; 4, 11 = 12, 50 = 11, 51 = 21, 52 = 31, \\ 53 = 32, 54 = 21, 55 = 10, 56 = Y, 57 = V, 58 = W, 59 = 23; 5, 10 = 12; \\ 5, 11 = 01, 60 = 00, 61 = 10, 62 = 20, 63 = 30, 64 = 22, 65 = 11, 66 = 00, \\ 67 = 10, 68 = 20, 69 = 30; 6, 10 = 22; 6, 11 = 11, 70 = 10, 71 = Y, 72 = V, \\ 73 = W, 74 = 23, 75 = 12, 76 = 01, 77 = 11, 78 = 21, 79 = 31; 7, 10 = 32; \\ 7, 11 = 21, 80 = 20, 81 = V, 82 = X, 83 = Z, 84 = 24, 85 = 13, 86 = 02, \\ 87 = 12, 88 = 22, 89 = 32; 8, 10 = 42; 8, 11 = 31, 90 = 30, 91 = W, 92 = Z, \\ 93 = 09, 94 = 19, 95 = 14, 96 = 03, 97 = 13, 98 = 23, 99 = 30; 9, 10 = 31; \\ 9, 11 = 32; 10, 0 = 22; 10, 1 = 23; 10, 2 = 24; 10, 3 = 19; 10, 4 = 08; \\ 10, 5 = 15; 10, 6 = 04; 10, 7 = 14; 10, 8 = 24; 10, 9 = W; 10, 10 = 20; \\ 10, 11 = 21; 11, 0 = 11; 11, 1 = 12; 11, 2 = 13; 11, 3 = 14; 11, 4 = 15; \\ 11, 5 = 07; 11, 6 = 05; 11, 7 = 15; 11, 8 = 19; 11, 9 = Z; 11, 10 = V; \\ 11, 11 = 10; \end{aligned}$$

* We shall see that in some cases the ambiguity of the sign of c may be removed by choice of the primitive root g , while in the remaining case the sign is fixed by the condition that the (ij) be integers.

where $X = (0, 10)$, $Y = (0, 11)$, $Z = (1, 10)$, $V = (1, 11)$, $W = (2, 11)$. Hence the 144 numbers (ij) reduce to 31:

$$(139) \quad 00, \dots, 09, 10, \dots, 15, 19, 20, \dots, 24, 30, 31, 32, 42, \\ X, Y, Z, V, W.$$

Write $[ij]$ for $(ij)_6$ for the ten $[ij]$ in

$$(140) \quad \begin{aligned} (01) &= [01] - (07) - 2(10), & 02 &= [02] - (08) - 2(20), \\ (03) &= [03] - (09) - 2(30), & 04 &= [04] - X - 2(22), \\ (05) &= [05] - Y - 2(11), & 06 &= [00] - 3(00), \\ (12) &= [12] - V - 15 - 21, & (13) &= [13] - W - 19 - 31, \\ (14) &= [14] - Z - (23) - (32), & (42) &= [24] - 3(24), \end{aligned}$$

which follow from (77). Then (17) reduce to (87) with f replaced by $2f$, and

$$(141) \quad \begin{aligned} (07) &= a + f + Y + W - 10 + 11 - 15 + 21 + 23 + 31 + 32, \\ (08) &= b - \frac{1}{2}(f + 1) + X + Z - W - 00 - 10 - 11 - 15 - 19 \\ &\quad - 2(20) - 21 - 2(24) - 30 + 32, \\ (142) \quad (22) &= \frac{1}{2}(f - 1) - 00 - 10 - 11 - 20 - 30, \\ (143) \quad a &= -[05] - [12] - [13] - [14], \quad b = [02] + [12] + [13] + [24]. \end{aligned}$$

By (27), $R(33) = h + 2n\beta^3$, $p = h^2 + 4n^2$, where

$$\begin{aligned} h &= -00 - 01 + 3(02) - 03 - 04 - 05 + 06 - 07 - 08 - 09 + 3X - Y \\ &\quad + 2\{10 + 11 - 12 - 13 + 14 + 15 + 19 - Z - V - 20 + 21 - 22 \\ &\quad + 23 - 24 + W + 30 - 31 - 32 + 42\}, \\ n &= -01 + 03 - 05 + 07 - 09 + 2(12) - 2(13) + 2(31) - 2(32) \\ &\quad + Y + 2Z - 2V. \end{aligned}$$

In $p = 12f + 1 = 4F + 1$, $F = 3f$ is odd. By § 12,

$$(144) \quad p = x^2 + 4y^2, \quad x \equiv 1 \pmod{4}, \quad 16(02)_4 = p + 1 - 6x, \quad y = (03)_4 - (01)_4.$$

By (138) we find that $n = y$, and by (140),

$$(145) \quad \frac{1}{3}(02)_4 = \frac{1}{3}\{[00] + 3[02] + 2[24]\} - (00) - (08) - 2(20) - 2(24) + X.$$

To h add (17) for $k = 0$ and eliminate the left members of (140) and (142). We get

$$\begin{aligned} h &\equiv -1 + 2\{[00] - [12] - [13] + [14] + [24]\} \pmod{4} \\ &\equiv -1 + 2\{[03] + [13] + [14] + 1\} \equiv -1 + 2(f + 1) \equiv -1, \end{aligned}$$

by (88) and (87₄). Hence $h = -x$ and

$$(146) \quad R(33) = -x + 2y\beta^3.$$

From the values in (27) we eliminate the left members of (140) and get

$$(147) \quad \begin{aligned} R(15) &= z + \beta^3 w, \quad R(17) = \rho + \beta^2 \sigma, \quad R(11) = H + \beta G + \beta^2 C + \beta^3 D, \\ R(12) &= R + \beta S + \beta^2 T + \beta^3 U, \quad R(13) = J + \beta K + \beta^2 P + \beta^3 Q; \end{aligned}$$

$$\begin{aligned} z &= z' - 4(00) + 2(08) + 4(10) + 4(11) - 2(15) - 2(19) - 2(20) \\ &\quad - 2(21) - 6(22) + 4(24) + 4(30) + 2(32) - 2X + 2Z - 2W, \\ z' &= [00] - [01] - [03] + 2[04] - [05] + [12] + [13] - [14] - [24], \\ w &= w' + 2\{07 + 2(09) + 10 + 11 - 15 - 2(19) - 21 - 3(23) + 2(30) \\ &\quad - 31 - 32 + Y - 2Z - 2V + W\}, \\ w' &= -[01] - 2[03] - [05] + 2[12] + [13] + 3[14], \\ \rho &= \rho' + 4[-00 + 10 - 15 + 19 - 21 - 2(24) - 30 - X + W], \\ \rho' &= [00] - [01] + [03] + [04] + 2[12] - 2[13] + 2[24], \\ \sigma &= \sigma' + 4\{08 - 10 + 11 - 19 + 2(20) - 32 + X - Z - W\}, \\ \sigma' &= [01] - 3[02] - [04] - [05] + 2[13] + 2[14], \\ H &= H' + 4\{-00 - 10 + 15 - 19 + 21 - 2(24) + 30 - X - W\}, \\ H' &= [00] + [01] - [03] + [04] - 2[12] + 2[13] + 2[24], \\ G &= 2[05] - 2[01] - 4[13] + 4\{07 + 10 - 11 + 19 + 2(31) + 32 \\ &\quad - Y - Z + W\}, \\ C &= C' + 4\{08 + 10 - 11 + 19 + 2(20) + 32 + X + Z + W\}, \\ C' &= -[01] - 3[02] - [04] + [05] - 2[13] - 2[14], \\ D &= -2[03] - 2[05] - 4[12] \\ &\quad + 4\{09 + 11 + 15 + 21 + 30 - 32 + Y + Z + 2V\}, \\ E &= E' + 2\{08 - 2(00) + 2(11) + 15 - 20 + 21 - 2(24) - 2(30) + 32 + Z\}, \\ E' &= -[00] + [03] - [05] - [12] - [14] + [24], \\ S &= S' + 2\{Y - 2Z - W - 07 - 10 + 11 + 2(19) - 3(23) + 31 - 32\}, \\ S' &= [01] - [05] - [13] + 3[14], \\ T &= T' + 2\{-08 + 2(10) - 2(11) - 19 + 20 - 3(22) - 32 - X - Z - W\}, \\ T' &= -[01] + 2[04] + [05] + [13] + [14], \\ U &= U' + 2\{07 - 2(09) + 10 + 15 - 2(19) + 21 - 2(30) - 31 + 2V + W\}, \\ U' &= -[01] + 2[03] - 2[12] + [13], \\ J &= J' + 2\{-2(00) - 08 + 2(11) + 15 + 20 \\ &\quad + 21 + 2(24) - 2(30) + 32 + Z\}, \\ K &= K' + 2\{07 + 10 - 11 - 2(15) + 2(21) - 23 - 31 - 32 - Y - W\}, \\ P &= P' + 2\{08 + 2(10) - 2(11) - 19 - 20 - 3(22) - 32 + X - Z - W\}, \\ Q &= Q' + 2\{2Y - W + 07 + 10 + 2(11) + 15 - 21 + 2(23) - 31 + 2(32)\}, \\ J' &= [00] + [03] - [05] - [12] - [14] - [24], \\ K' &= [05] - [01] + [13] + [14], \\ P' &= -[01] - 2[04] + [05] + [13] + [14], \\ Q' &= -[01] - 2[05] + [13] - 2[14]. \end{aligned}$$

The ten $[ij]$ were found in § 18 and are here regarded as known. The 31 numbers (139) are connected by the 10 + 3 equations (140)-(142), the 16 whose left members are z, \dots, Q , and the two final equations (144), amplified by (145) and (138). These 31 linear equations uniquely determine the 31 numbers (139) and hence all 144 of the (ij) .

We seek especially (00) and hence (06) by (140). We shall find 00, 24 and 30 simultaneously by three linear equations:

$$\begin{aligned}
 2\rho + \sigma + 2P + 4J &= s + 2P' + 4J' + 6f - 6 - 36(00) - 36(30), \\
 2\rho + \sigma + 4R + 2T - 2z &= s + 4R' + 2T' - 2z' + 4b - 2f - 2 \\
 (148) \quad &+ 12(00) - 48(24) - 36(30), \\
 2\rho + \sigma + 2H + C + \frac{8}{3}(02) &= s + 2H' + C' + \frac{8}{3}\{[00] + 3[02] \\
 &+ 2[24]\} - 24(00) - 48(24), \\
 s &= 2[00] - [01] - 3[02] + 2[03] + [04] \\
 &- [05] + 4[12] - 2[13] + 2[14] + 4[24].
 \end{aligned}$$

I. Let 2 be a cubic residue of p . Then m is an odd multiple of 3, and $\beta^{2m} = -1$.

In (148), insert the values (91) of the $[ij]$, and solve. We get

$$\begin{aligned}
 144(00) &= p - 23 - 20A + 2x - 2\tau + 2\phi, \quad \tau = 2\rho + \sigma + 2P + 4J, \\
 (149) \quad 144(30) &= p - 11 + 4A - 2x - 2\tau - 2\phi, \quad \phi = 4R + 2T - 2z - 2H - C, \\
 144(24) &= p + 1 - 2A + 2x + \tau - \phi - 3(2\rho + \sigma) - 6H - 3C.
 \end{aligned}$$

By I. of § 18, $L = E = -2A$, $F = 2B = 3M$. By (137), $R(22) = R(44) = -A - B + 2\beta^2 B$. By (130), $2\rho + \sigma = 2A$, $G = C = 0$, $H = z$, $D = w$.

I₁. Let 3 be a biquadratic residue of p . Then $k = +1$ in (132). By also (135), (136),

$$\begin{aligned}
 R(15) &= -R(33), \quad R(13) = \pm \beta^3 R(15), \quad R(12) = \mp \beta^3 R(22), \\
 z = x, \quad w &= -2y, \quad K = P = 0, \quad J = \pm 2y, \quad Q = \pm x, \quad R = T = 0, \\
 S &= \pm 2B, \quad U = \pm (A - B).
 \end{aligned}$$

The upper signs are replaced by the lower when g is replaced by g^r , $r \equiv -1 \pmod{12}$. By Theorem 10 we see that $x, z, A, E, L, K, G, S, 00, 30, 24$ are unaltered; $y, w, B, F, M, P, C, T, \sigma$ are changed in sign; and, if J becomes J_1 , then $J_1 = J + P$, $Q_1 = -Q - K$, $H_1 = H + C$, $D_1 = -D - G$, $R_1 = R + T$, $U_1 = -U - S$, $\rho_1 = \rho + \sigma$. We get

$$\begin{aligned}
 (150) \quad 144(00) &= p - 23 - 24A - 6x \mp 16y, \quad 144(30) = p - 11 + 6x \mp 16y, \\
 144(24) &= p + 1 - 6A \pm 8y.
 \end{aligned}$$

I_2 . Let 3 be a biquadratic non-residue of p . Then $k = -1$, $z = -x$, $w = 2y$, $K = P = S = U = 0$, $J = \pm x$, $Q = \mp 2y$, $R = \pm (A + B)$, $T = \mp 2B$,

$$(151) \quad \begin{aligned} 144(00) &= p - 23 - 24A + 10x \pm 8(A - x), \\ 144(30) &= p - 11 - 10x \mp 8(A + x), \\ 144(24) &= p + 1 - 6A + 4x \mp 4(A - x). \end{aligned}$$

The signs are not affected by the choice of the primitive root g , but are determined by the fact that the right members shall be integers. The upper signs hold if $p = 157$, the lower if $p = 397$ or 997 (the only p 's < 1000).

For $m \equiv 1$ or $4 \pmod{6}$, $E, \dots, M, (ij)$ are given by III of § 18. Then (148) give

$$(152) \quad \begin{aligned} 144(00) &= p - 23 + 2x - 8A + 6B + 2(v - u), \\ 144(30) &= p - 11 - 2x + 4A - 6B - 2(v + u), \\ 144(24) &= p + 1 + 2x + 10A + 6B - v + u - 3(2\rho + \sigma + 2H + C), \\ u &= 2\rho + \sigma + 2P + 4J, \quad v = 4R + 2T - 2z - 2H - C. \end{aligned}$$

II₁. Let $2m \equiv 2 \pmod{12}$, $k = -1$. Then (130)-(137) give $C = z = -x$, $w = 2y$, $H = x$, $G = -2y$, $D = 0$, $2\rho + \sigma = -A + 3B$, $J = \pm x$, $Q = \mp 2y$, $R = \pm (A + B)$, $T = \mp 2B$, $K = P = S = U = 0$.

$$(153) \quad \begin{aligned} 144(00) &= p - 23 + 4x - 6A \pm 8(A - x), \\ 144(24) &= p + 1 - 2x + 12A \mp 4(A - x), \\ 144(30) &= p - 11 - 4x + 6A - 12B \mp 8(A + x). \end{aligned}$$

II₂. Let $2m \equiv 2 \pmod{12}$, $k = 1$. Then

$z = C = x$, $H = -x$, $Q = \pm x$, $G = 2y$, $w = -2y$, $J = \pm 2y$, $D = K = P = R = T = 0$, $S = \pm 2B$, $U = \pm (A - B)$, $2\rho + \sigma = -A + 3B$,

$$(154) \quad \begin{aligned} 144(00) &= p - 23 - 6A \mp 16y, \quad 144(24) = p + 1 + 6x + 12A \pm 8y, \\ 144(30) &= p - 11 + 6A - 12B \mp 16y. \end{aligned}$$

We may take the upper signs. For, if g be replaced by a new primitive root g^r , $r \equiv 7 \pmod{12}$, k is unchanged, $2m$ is unaltered modulo 12, y is changed in sign, while

$$(155) \quad (00), (30), (24), x, A$$

and B are unaltered.

If $r \equiv -1 \pmod{12}$, we saw under I_1 that (155) are unaltered while

y and B are changed in sign. If $r \equiv 5 \pmod{12}$, (155) and y are unaltered, while B is changed in sign. This proves

III₁. If $2m \equiv 10 \pmod{12}$, $k = -1$, change the sign of B in II₁.

III₂. If $2m \equiv 10 \pmod{12}$, $k = 1$, change the sign of B in II₂. The upper signs hold when g is chosen properly. There are no further cases with f odd, since m is then odd.

28. Case $e = 12$, f even. We replace (144), (145) by

$$\begin{aligned} p &= x^2 + 4y^2, \quad x \equiv 1 \pmod{4}, \quad 16(00)_4 = p - 11 - 6x, \quad y = (01)_4 - (03)_4, \\ \frac{1}{3}(00)_4 &= \frac{1}{3}\{[00] + 3[02] + 2[24]\} \\ &\quad - (02) + (04) - (06) - 2(26) - 2(2, 10). \end{aligned}$$

The further formulas for f odd hold here if we change the signs of the expressions for R, S, T, U (and hence of R', \dots), replace (k, h) by $(k, h + 6)$, and change the constant terms as follows: to $-\frac{1}{2}f$ in (141₂), to $\frac{1}{2}f$ in (142), suppress -6 and -2 from the first and second equations (148), replace $(02)_4$ by $(00)_4$ in the third.

I. Let 2 be a cubic residue of p . Then $\beta^{2m} = +1$. Now $2\rho + \sigma = -2A$. Change the constant terms in (149) to $-11, 1, 1$.

I₁. Let 3 be a biquadratic residue of p . Then $k = 1$,

$$\begin{aligned} H = z = -x, \quad D = w = 2y, \quad J = \pm x, \quad Q = \mp 2y, \quad R = \pm (A + B), \\ T = \mp 2B, \quad G = C = K = P = S = U = 0, \end{aligned}$$

$$\begin{aligned} (156) \quad 144(06) &= p - 11 + 10x - 16A \mp 8(A + x), \\ 144(36) &= p + 1 - 10x + 8A \pm 8(A - x), \\ 144(2, 10) &= p + 1 + 4x + 2A \pm 4(A + x). \end{aligned}$$

I₂. Let 3 be a biquadratic non-residue of p . Then $k = -1$,

$$\begin{aligned} H = z = x, \quad D = w = -2y, \quad J = \pm 2y, \quad Q = \pm x, \quad S = \pm 2B, \quad U = \pm (A - B), \\ G = C = K = P = R = T = 0, \end{aligned}$$

$$\begin{aligned} (157) \quad 144(06) &= p - 11 - 6x - 16A \mp 16y, \quad 144(2, 10) = p + 1 + 2A \pm 8y, \\ 144(36) &= p + 1 + 6x + 8A \mp 16y. \end{aligned}$$

For $m \equiv 4 \pmod{6}$, the products of (06) , (36) , $(2, 10)$ by 144 are given by (152) with the constant terms replaced by $-11, 1, 1$ and

II₁. Let $2m \equiv 8 \pmod{12}$, $k = -1$. Then

$$C = z = x, \quad w = -2y, \quad G = 2y, \quad H = -x, \quad J = \pm 2y, \quad 2\rho + \sigma = A - 3B, \\ Q = \pm x, \quad S = \pm 2B, \quad U = \pm (A - B), \quad D = K = P = R = T = 0,$$

$$(158) \quad \begin{aligned} 144(06) &= p - 11 - 10A + 12B \mp 16y, \quad 144(36) = p + 1 + 2A \mp 16y, \\ 144(2, 10) &= p + 1 + 6x + 8A + 12B \pm 8y. \end{aligned}$$

The sign of y is changed when g is replaced by g^r , $r \equiv 7 \pmod{12}$.

II₂. Let $2m \equiv 8 \pmod{12}$, $k = +1$. Then

$$C = z = -x, \quad w = 2y, \quad H = x, \quad G = -2y, \quad J = \pm x, \quad 2\rho + \sigma = A - 3B, \\ Q = \mp 2y, \quad R = \pm (A + B), \quad T = \mp 2B, \quad D = K = P = S = U = 0,$$

$$(159) \quad \begin{aligned} 144(06) &= p - 11 + 4x - 10A + 12B \mp 8(A + x), \\ 144(36) &= p + 1 - 4x + 2A \pm 8(A - x), \\ 144(2, 10) &= p + 1 - 2x + 8A + 12B \pm 4(A + x). \end{aligned}$$

III. Let $2m \equiv 4 \pmod{12}$. When g is replaced by g^r , $r \equiv -1 \pmod{12}$, (06), (36), (2, 10), A , x remain unaltered, while B and y are changed in sign. Making the latter change in II₁ and II₂, we obtain the present values of 144(06), etc.

THE UNIVERSITY OF CHICAGO.

SPINORS IN n DIMENSIONS.

By RICHARD BRAUER AND HERMANN WEYL.

Introduction and Summary. Let \mathfrak{d}_n be the group of orthogonal transformations o :

$$(1) \quad x_i \rightarrow \sum_{k=1}^n o(ik) x_k \quad (i = 1, 2, \dots, n)$$

of the n -dimensional space, and \mathfrak{d}_n^+ the subgroup of proper transformations, having determinant $+1$ and not -1 . We shall first operate within the continuum of all complex numbers, whereas the particular conditions prevailing under restriction to real variables will be studied only at the end of the paper (§§ 8 and 10). A given representation $\Gamma: o \rightarrow G(o)$ of degree N defines a certain kind of "covariant quantities": a quantity characterized by N numbers a_1, \dots, a_N relative to an arbitrary Cartesian coördinate system in the underlying n -dimensional Euclidean space will be called a *quantity of kind Γ* , provided the components a_K experience the linear transformation $G(o)$ under the influence of the coördinate transformation o . The quantity is called *primitive* if the representation is irreducible. The proposition that every representation breaks up into irreducible parts, states that the most general kind of quantities is obtained by juxtaposition of several independent primitive quantities.

By a *tensor of rank f* we shall mean here what usually is called a skew-symmetric tensor: a skew-symmetric function $\alpha(i_1 \dots i_f)$ of f indices ranging independently from 1 to n which transforms according to the law

$$\alpha(i_1 \dots i_f) \rightarrow \sum_{k_1, \dots, k_f=1}^n o(i_1 k_1) \dots o(i_f k_f) \cdot \alpha(k_1 \dots k_f)$$

under the influence of the rotation o . The tensors of rank f form the substratum of a representation Γ_f of degree $\binom{n}{f}$.

We often have to distinguish between even and odd dimensionality, and we shall accordingly put $n = 2\nu$ or $n = 2\nu + 1$. Let us use the notation $\nu = \langle n \rangle$ and in passing notice the congruence

$$\frac{1}{2}n(n-1) \equiv \langle n \rangle \pmod{2}.$$

E. Cartan developed a general method of constructing irreducible representations of \mathfrak{d}_n (or any other semi-simple group) by considering the in-

finitesimal operations, and he found † as the building stones of the whole edifice the tensor representations Γ_i together with *one further double-valued representation* $\Delta : o \rightarrow S(o)$ of degree 2^v . The quantities of kind Δ are called *spinors*. In the four-dimensional world this kind of quantities has come to its due honors by Dirac's theory of the spinning electron. Cartan, according to his standpoint, states the transformation law $S(o)$ of spinors only for the infinitesimal rotations o . Here we shall give a simple finite description of the representation Δ and shall derive from it by the simplest algebraic means the main properties of the spinors. One will be able to judge by this theory to what extent recent investigations about spinor calculus reveal those essential features that stay unchanged for higher dimensions. One of the chief results will be that Dirac's equations of the motion of an electron and the expression for the electric current are uniquely determined even in the case of arbitrary dimensionality.

Our investigation will be arranged as follows: we start (§ 2) with a certain associative algebra Π of order 2^{2v} which proves to be a complete matrix algebra in 2^v dimensions, and leads to the desired definition of Δ (§ 3). We shall first get Δ as a collineation representation such that only the ratios of the spinor components have a meaning. In the case of even dimensionality $n = 2^v$ we shall prove (§ 3) that the product $\Delta \times \check{\Delta}$ of Δ by the contragredient representation $\check{\Delta}$ splits up according to the equivalence:

$$\Delta \times \check{\Delta} \sim \Gamma_0 + \Gamma_1 + \Gamma_2 + \cdots + \Gamma_n,$$

whereas in the odd case

$$\Delta \times \check{\Delta} \sim \Gamma_0 + \Gamma_2 + \Gamma_4 + \cdots + \Gamma_{n-1}$$

(§ 5). The collineation representation Δ can be normalized so as to give an ordinary, though double-valued representation Δ satisfying the equivalence $\check{\Delta} \sim \Delta$ (§§ 4, 5). If one restricts oneself to the proper orthogonal transformations in a space of even dimensionality, Δ splits up into two representations Δ^+ and Δ^- each of degree 2^{v-1} (§ 6). The four products of the type $\Delta \times \check{\Delta}$ will be determined individually for $\Delta = \Delta^+$ or Δ^- , and so will the equivalences of type $\check{\Delta} \sim \Delta$. The transition from our finite to Cartan's infinitesimal description can be easily performed (§ 7). In considering real transformations only, the differences of the inertial index have to be taken into account (§ 8); it will be proved that $\check{\Delta}$ is equivalent to Δ again—but for a sign the determination of which is of peculiar interest and closely related

† *Bulletin Société Mathématique de France*, vol. 41 (1913), p. 53. Compare also Weyl, *Mathematische Zeitschrift*, vol. 24 (1926), p. 342.

to the inertial index. Irreducibility and equivalence of the occurring representations will be ascertained in § 9, and the relation to physics will be discussed in § 10. In parts of the investigation we must have recourse to the law of duality of tensors and tensor representations Γ_f as formulated in the preliminary § 1. The last section (§ 11) is devoted to the demonstration of a well-known fundamental proposition concerning the automorphisms of the complete matrix algebra, a proposition indispensable for the definition of Δ .

1. *Duality of tensors.* Γ_n is the representation of degree 1 of the full rotation group \mathfrak{d}_n associating the signature $\sigma(o)$ with the rotation $o: \sigma(o) = +1$ for the proper, $\sigma(o) = -1$ for the improper rotations. Any representation $\Gamma: o \rightarrow G(o)$ gives rise to another representation $\sigma\Gamma: o \rightarrow \sigma(o)G(o)$, coinciding with Γ under restriction to \mathfrak{d}_n^+ .

The equation

$$(2) \quad \alpha^*(i'_1 \cdots i'_{n-f}) = \alpha(i_1 \cdots i_f)$$

in which $i_1 \cdots i_f i'_1 \cdots i'_{n-f}$ denotes any even permutation of the figures from 1 to n , associates a tensor α^* of rank $n - f$ with every tensor α of rank f . This relation is invariant with respect to proper orthogonal transformations. Thus the *law of duality* $\Gamma_{n-f} \sim \Gamma_f$ prevails for the tensor representations Γ_f of \mathfrak{d}_n^+ . When taking the improper orthogonal transformations into consideration it is to be replaced by

$$\Gamma_{n-f} \sim \sigma\Gamma_f.$$

In the case of an even number of dimensions $n = 2\nu$, the representation Γ_ν deserves particular attention. It satisfies the equivalence $\sigma\Gamma_\nu \sim \Gamma_\nu$. (2) or rather

$$(3) \quad \alpha^*(i'_1 \cdots i'_\nu) = i^\nu \cdot \alpha(i_1 \cdots i_\nu)$$

now establishes a transformation $\alpha \rightarrow \alpha^*$ of the space of the tensors of rank ν upon itself. We added the factor i^ν in order to make this transformation involutorial: $\alpha^{**} = \alpha$; for if $i_1 \cdots i_\nu i'_1 \cdots i'_\nu$ is an even permutation, $i'_1 \cdots i'_\nu i_1 \cdots i_\nu$ has the character $(-1)^\nu$. We may distinguish between positive and negative tensors of rank ν according as $\alpha^* = \alpha$ or $\alpha^* = -\alpha$. Any tensor of rank ν can be decomposed in a unique manner into a positive and a negative part:

$$\alpha = \frac{1}{2}(\alpha + \alpha^*) + \frac{1}{2}(\alpha - \alpha^*).$$

Hence, as a representation of the group $\mathfrak{d}_{2\nu}^+$, Γ_ν splits up into two representations $\Gamma_\nu^+ + \Gamma_\nu^-$ of half the degree.

2. *The algebra II.* Our procedure is exactly the same as followed by

Dirac in his classical paper on the spinning electron.[†] We introduce n quantities p_i which turn the fundamental quadratic form into the square of a linear form:

$$(4) \quad x_1^2 + \cdots + x_n^2 = (p_1 x_1 + \cdots + p_n x_n)^2.$$

For this purpose we must have

$$(5) \quad p_i^2 = 1, \quad p_k p_i = -p_i p_k \quad (k \neq i).$$

The quantities p_i engender an algebra consisting of all linear combinations of the 2^n units

$$(6) \quad e_{a_1 \dots a_n} = p_1^{a_1} \cdots p_n^{a_n} \quad (\alpha_1, \dots, \alpha_n \text{ integers mod } 2).$$

The recipe for multiplication of the units reads, according to (5):

$$e_{a_1 \dots a_n} \cdot e_{\beta_1 \dots \beta_n} = (-1)^\delta \cdot e_{\gamma_1 \dots \gamma_n}; \quad \gamma_i = \alpha_i + \beta_i, \quad \delta = \sum_{i > k} \alpha_i \beta_k.$$

One easily convinces oneself that this rule of multiplication is associative.

One may write the most general quantity a of our algebra in the form

$$(7) \quad a = \cdots + (1/f!) \sum_{(i_1, \dots, i_f)} \alpha(i_1 \cdots i_f) p_{i_1} \cdots p_{i_f} + \cdots \quad (f = 0, 1, \dots, n),$$

splitting a into parts according to the number f of the different factors p . Since the product of f different p 's like $p_{i_1} \cdots p_{i_f}$ is skew-symmetric with respect to the indices $i_1 \cdots i_f$, one will choose the coefficients $\alpha(i_1 \cdots i_f)$ in (7) also skew-symmetric; one is then allowed to extend the sum Σ in (7) over the indices i_1, \dots, i_f independently from 1 to n . Consequently the quantity a is equivalent to a "tensor set" consisting of $n+1$ tensors, one of each of the ranks $0, 1, \dots, f, \dots, n$. The addition of two tensor sets and the multiplication of a set by a number has the trivial significance within the algebra II. But how are we to express the multiplication of two tensor sets a and b ? It suffices to describe the case of an a containing merely one tensor α of rank f , and a b containing merely one tensor β of rank g (whereas the other parts vanish). The product splits into different parts according to the number r of coincidences among the indices of α and β . As

$$\begin{aligned} p_{i_1} \cdots p_{i_{f-r}} p_{i_1} \cdots p_{i_r} p_{i_1} \cdots p_{i_r} p_{k_1} \cdots p_{k_{g-r}} \\ = (-1)^{r(r-1)/2} p_{i_1} \cdots p_{i_{f-r}} p_{k_1} \cdots p_{k_{g-r}}, \end{aligned}$$

one gets as part r of the product essentially the "contraction"

[†] *Proceedings of the Royal Society (A)*, vol. 117 (1927), p. 610; vol. 118 (1928), p. 351.

$$(8) \quad \gamma(i_1 \cdots i_{f-r} k_1 \cdots k_{g-r}) = \sum_{(l_1, \dots, l_r)} \alpha(i_1 \cdots i_{f-r} l_1 \cdots l_r) \cdot \beta(l_1 \cdots l_r k_1 \cdots k_{g-r}).$$

This process, however, has to be followed by "alternation," i. e. alternating summation over all permutations of the $f + g - 2r$ indices in γ . Since γ is already skew-symmetric with respect to the $f - r$ indices i and the $g - r$ indices k , it is sufficient to extend an alternating sum over all "mixtures" of the indices $i_1 \cdots i_{f-r}$ with the indices $k_1 \cdots k_{g-r}$. This will be indicated by the symbol M . By taking into consideration the factor $1/f!$ attached to the f -th term in (7) and the several distributions of the r equal indices $l_1 \cdots l_r$ among the indices of α and β , one gets finally the result: The "product" of the two tensors α and β is a tensor set in which only tensors of rank $f + g - 2r$ appear; the integer r is limited by the bounds

$$r \geq 0, \quad 2r \leq f + g - n, \quad r \leq f, \quad r \leq g.$$

The part r is given by

$$(-1)^{\langle r \rangle} (1/r!) \cdot M \gamma(i_1 \cdots i_{f-r} k_1 \cdots k_{g-r})$$

where γ denotes the contraction (8).—We are not so much interested in the exact description of this process of multiplication as in the fact *that it is orthogonally invariant*.

3. *Spinors in a space of even dimensionality.* In this section we suppose $n = 2\nu$ to be even. The algebra Π is known to the quantum theorist from the process of "superquantizing" that allows the passage from the theory of a single particle to the theory of an undetermined number of equal particles subjected to the Fermi statistics. This connection at once yields a definite representation $p_i \rightarrow P_i$ by matrices P_i of order 2^ν . Into its description enter the two-rowed matrices

$$1 = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}, \quad 1' = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix}, \quad P = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}, \quad Q = \begin{vmatrix} 0 & i \\ -i & 0 \end{vmatrix}.$$

The two rows and columns will be distinguished from each other by the signs $+$ and $-$. $1', P, Q$ anticommute with each other; their squares are $= 1$. Besides $p_1, \dots, p_{2\nu}$ we sometimes use the notation $p_1, \dots, p_\nu, q_1, \dots, q_\nu$. The representation then is given by

$$(9) \quad \begin{aligned} p_\alpha &\rightarrow P_\alpha = 1' \times \cdots \times 1' \times P \times 1 \times \cdots \times 1, \\ q_\alpha &\rightarrow Q_\alpha = 1' \times \cdots \times 1' \times Q \times 1 \times \cdots \times 1. \end{aligned} \quad (\alpha = 1, \dots, \nu).$$

On the right side we have ν factors; the factors P, Q respectively, occur at the α -th place. The rows and columns of our matrices or the coördinates x_A in

the 2^v -dimensional representation space, according to the notation introduced, are distinguished from each other by a combination of signs $(\sigma_1, \sigma_2, \dots, \sigma_v)$, $(\sigma_a = \pm)$. One verifies at once that the desired rules prevail:

$$(10) \quad P_i^2 = 1, \quad P_k P_i = -P_i P_k \quad (i \neq k).$$

In this manner we have established a definite representation $x \rightarrow X$ of degree 2^v for the algebra Π . We maintain that *all matrices X appear here as images of elements x of the algebra*. As the algebra Π is of the same order $2^{2v} = (2^v)^2$ as the algebra consisting of all matrices in the 2^v -dimensional space, the relation $x \rightleftharpoons X$ is a one-to-one isomorphic mapping of Π upon the complete matrix algebra of the 2^v -dimensional "spin space": the algebra Π is isomorphic to the complete matrix algebra in spin space. In order to prove our statement, let us compute the matrix U_a representing $u_a = ip_a q_a$:

$$(11) \quad U_a = iP_a Q_a = 1 \times \dots \times 1 \times 1' \times 1 \times \dots \times 1$$

and then

$$(11') \quad U_1 \dots U_{a-1} P_a = 1 \times \dots \times 1 \times P \times 1 \times \dots \times 1$$

together with $U_1 \dots U_{a-1} Q_a$. (The factors different from 1 occur at the α -th place.) Thus the following elements

$$\begin{aligned} \frac{1}{2}(1 + u_a) &= z_a^{++}, & \frac{1}{2}u_1 \dots u_{a-1}(p_a - iq_a) &= z_a^{+-}, \\ \frac{1}{2}u_1 \dots u_{a-1}(p_a + iq_a) &= z_a^{-+}, & \frac{1}{2}(1 - u_a) &= z_a^{--} \end{aligned}$$

are represented by products similar to (11) but containing one of the matrices

$$\begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix}, \quad \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix}, \quad \begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix}, \quad \begin{vmatrix} 0 & 0 \\ 0 & 1 \end{vmatrix}$$

at the α -th place. Consequently the image of the element $\prod_{a=1}^v (z_a^{\sigma_a \tau_a})$ is the matrix containing a term different from 0, namely 1, only at the crossing point of the row $\sigma_1 \dots \sigma_v$ with the column $\tau_1 \dots \tau_v$ ($\sigma_a = \pm, \tau_a = \pm$).

We are now in a position to establish the connection with the rotations $o = \|o(ik)\|$ in the n -dimensional space (Method A). We change, by means of the orthogonal matrix $o(ik)$

$$(12) \quad P_i \rightarrow P_i^* = \sum_{k=1}^n o(ki)P_k, \quad P_i = \sum_{k=1}^n o(ik)P_k^*$$

and we observe at once that the new P_i^* , like the old ones, satisfy the relations (10). Consequently $p_i \rightarrow P_i^*$ defines a new representation of our algebra Π . Since the full matrix algebra, however, allows only inner automorphisms,[†]

[†] See the proof in § 11.

this representation has to be equivalent to the original one; that is, there exists a non-singular matrix $S(o)$ such that

$$(13) \quad P^*_i = S(o)P_i S(o)^{-1} \quad (i = 1, 2, \dots, n).$$

$S(o)$ is determined by this equation but for a numerical factor, the "gauge factor": $S(o)$ is to be interpreted in the "homogeneous" sense, not as an affine transformation of the 2^n -dimensional vector space, but as a collineation of the projective space consisting of its rays. After fixing the gauge factors for two rotations o, o' and their product $o'o$ in an arbitrary manner, we necessarily have a relation like

$$(14) \quad S(o'o) = c \cdot S(o')S(o).$$

Consequently we are dealing with a *collineation representation* of degree 2^n of the rotation group, the so-called *spin representation* $\Delta: o \rightarrow S(o)$.

The same connection can be described as follows (Method B). Orthogonal transformation of the tensors of an arbitrary tensor set defines an automorphic mapping $x \rightarrow x^*$ of the algebra Π of the tensor sets upon itself. Such a mapping however, in the representation $x \rightarrow X$ of the tensor sets by matrices X of order 2^n , is necessarily displayed in the form

$$X \rightarrow X^* = SXS^{-1} \quad (S \text{ independent of } x).$$

Let us write down this equation in components: $X = \|x_{JK}\|$; it then reads

$$x^*_{JK} = \sum_{R,T} s_{JR} \check{s}_{KT} x_{RT}.$$

$\check{S} = \|\check{s}_{JK}\|$ is the matrix contragredient to S . Hence the components x_{JK} experience the transformation $S \times \check{S}$ and this proves the reduction

$$(15) \quad \Delta \times \check{\Delta} \sim \Gamma_0 + \Gamma_1 + \dots + \Gamma_n \sim \left\{ \begin{array}{c} \Gamma_0 + \Gamma_1 + \dots + \Gamma_{n-1} + \\ \sigma\Gamma_0 + \sigma\Gamma_1 + \dots + \sigma\Gamma_{n-1} \end{array} \right\} + (\Gamma_n \sim \sigma\Gamma_n).$$

The quantities $\{\psi^A\}$ and $\{\phi_A\}$ of the kind $\Delta, \check{\Delta}$ shall be called *covariant and contravariant spinors* respectively. Let us write the components ψ^A of a covariant spinor as a column and the components ϕ_A of a contravariant spinor as a row. Our last equation tells us that one is able to form by linear combination of the $(2^n)^2$ products $\phi_A \psi^B$: one scalar, one vector, one tensor of rank 2, etc. The scalar is, of course,

$$\phi\psi = \sum_A \phi_A \psi^A.$$

The vector has the components $\phi P_i \psi$. Indeed, in carrying out the transformation $\psi^* = S\psi$, $\phi^* = \phi S^{-1}$, one gets,

$$\phi^* P_i \psi^* = \phi S^{-1} P_i S \psi = \sum_{k=1}^n o(ik) \phi S^{-1} P_k^* S \psi = \sum_{k=1}^n o(ik) \phi P_k \psi.$$

The tensor of rank 2 has the components $\phi(P_i P_k) \psi$ [$i \neq k$]; etc. In this manner we are able to carry out the reduction (15) explicitly.

4. *Connection between covariant and contravariant spinors.* Let n be even as before. We propose to show that the representation Δ is equivalent to the representation Δ . For this purpose we observe that the relations (10) characteristic for the matrices P_i hold at the same time for the transposed matrices P'_i . According to the proposition on the automorphisms of our matrix algebra Π we already have had occasion to use, there must exist a definite non-singular matrix C such that

$$(16) \quad P'_i = C P_i C^{-1}$$

for all i . It is easy to write down C explicitly. For we have

$$P'_a = P_a, \quad Q'_a = -Q_a \quad (\alpha = 1, \dots, \nu).$$

But the product $p_1 \cdots p_\nu$ commutes with the p_a and anticommutes with the q_a , if ν is odd; if ν is even the situation is reversed. Hence one can take

$$c = p_1 \cdots p_\nu \quad \text{or} \quad = q_1 \cdots q_\nu$$

according as ν is odd or even. In this way one finds in both cases:

$$(17) \quad C = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \times \begin{vmatrix} 0 & i \\ -i & 0 \end{vmatrix} \times \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \times \cdots \quad (\nu \text{ factors})$$

and one verifies at once the relations (16).

Along with (12) we have

$$P'_i \rightarrow P^{*'}_i = \sum_k o(ki) P'_k.$$

This transition is expressed on the one hand in the form

$$P'_i \rightarrow S'(o)^{-1} P'_i S'(o) = \check{S}(o) P'_i \check{S}(o)^{-1}.$$

On the other hand the transformation of $P'_i = C P_i C^{-1}$ is obviously performed by means of $CS(o)C^{-1}$. Hence an equation like

$$CS(o)C^{-1} = \rho(o) \cdot \check{S}(o)$$

must hold where $\rho(o)$ is a numerical factor dependent on o . On multiplication of $S(o)$ by λ , $\check{S}(o)$ is multiplied by $1/\lambda$ and ρ is thus changed into $\rho\lambda^2$.

Hence we may dispose of the arbitrary gauge factor in S in such a way that ρ becomes $= 1$:

$$(18) \quad \check{S}(o) = CS(o)C^{-1}.$$

This has the effect that

$$(19) \quad (\det S)^2 = 1.$$

$S(o)$ is now uniquely determined *but for the sign*. After normalizing this sign for two rotations o, o' and the compound $o'o$ in an arbitrary manner, the composition factor c in (14) becomes $= \pm 1$; for the matrices $X = S(o'o)$ and $X = S(o')S(o)$ both satisfy the normalizing condition

$$\check{X} = CXC^{-1}.$$

Δ now is an ordinary, though double-valued representation instead of a collineation representation.

Equation (18) gives the explicit relation between the covariant and contravariant spinors: if C is the matrix $\|c_{AB}\|$ the substitution

$$\phi_A = \sum_B c_{AB} \psi^B.$$

changes the covariant spinor ψ into a contravariant spinor ϕ .

The "square" of the double-valued representation Δ is single-valued and is decomposed, according to formula

$$\Delta \times \Delta \sim \Gamma_0 + \Gamma_1 + \cdots + \Gamma_{n-1} + \Gamma_n$$

into the tensor representations Γ_f .

5. *Odd number of dimensions.* $n = 2\nu + 1$. To our quantities $p_1, \dots, p_{2\nu}$ a further one $p_{2\nu+1}$ has to be added, $p_{2\nu+1}^2 = 1$, which anticommutes with the previous p_i . The representation $p_i \rightarrow P_i$ ($i = 1, \dots, 2\nu$) can be extended by establishing the correspondence

$$p_n \rightarrow P_n = 1' \times 1' \times \cdots \times 1' \quad (n = 2\nu + 1).$$

Let ι be $= 1$ or i according as ν is even or odd. The product

$$(20) \quad u = \iota p_1 p_2 \cdots p_n$$

commutes with all quantities of the algebra and satisfies the equation $u^2 = 1$. In the representation just described u is represented by the matrix 1. There exists a second representation of the algebra:

$$(21) \quad p_i \rightarrow -P_i \quad (i = 1, 2, \dots, n)$$

in which $u \rightarrow -1$ and which thus proves to be inequivalent to the first one.

The order $2 \cdot (2^n)^2$ of the algebra Π this time is twice as large as the order of the algebra of all matrices X in the 2^n -dimensional spin-space. Our isomorphic mapping $x \rightarrow X$ therefore becomes a one-to-one correspondence only after reducing Π modulo $(1 - u)$; this is accomplished by adding the condition $u = 1$ to the defining equations (5). This new algebra may be realized as a subalgebra in Π in different manners; for instance, as the algebra of the quantities x satisfying the condition $x = ux$. It is more convenient to consider the *even* quantities in Π . Their basis consists of the products of an even number of p ; in (6) one has to add the restriction $\alpha_1 + \dots + \alpha_n \equiv 0 \pmod{2}$; the corresponding tensor sets contain tensors of even rank only. Any odd quantity may be written in the form ux where x is even. The arbitrary quantity $x + ux'$ of the algebra Π (x and x' even) is represented by the same matrix as the even quantity $x + x'$. Hence the correspondence $x \rightarrow X$ is a one-to-one correspondence within the algebra Π_e of the even quantities. The second representation (21) coincides with the first for the even quantities.

The procedure is now as above (Method A). Let $\|o(ik)\|$ be a proper orthogonal transformation. Then (12) yields a new representation of Π . By multiplication we get

$$U^* = iP^*_1 \cdot \dots \cdot P^*_n = \det[o(ik)] \cdot U = U.$$

Hence this representation like the original one associates the matrix $+1$ (and not -1) with u ; by means of $P_i \rightarrow P^*_i$ we thus map the algebra Π reduced modulo $(1 - u)$ isomorphically upon itself, and consequently an equation like

$$P^*_i = SP_i S^{-1}$$

holds. The representation $\Delta : o \rightarrow S(o)$ may be extended to the improper rotations by making the matrix $+1$ or -1 correspond to the reflection $x_i \rightarrow -x_i$ that commutes with all rotations. (Whether one chooses $+1$ or -1 does not make any difference here since the representation Δ is double-valued.)

(Method B). The orthogonal transformation o is an isomorphic mapping of the manifold of all even tensor sets upon itself. After representing this manifold by the algebra of all matrices X in 2^n dimensions in the manner described above, o appears as an automorphism $X \rightarrow X^*$ of the complete matrix algebra: $X^* = SX S^{-1}$. One gets $S(o)$ here at the same time for all proper and improper rotations o . Furthermore, we obtain the decomposition

$$(22) \quad \Delta \times \tilde{\Delta} \sim \Gamma_0 + \Gamma_2 + \dots + \Gamma_{2\nu} \sim \Gamma_0 + \sigma\Gamma_1 + \Gamma_2 + \sigma\Gamma_3 + \dots,$$

the last sum concluding with the term Γ_ν or $\sigma\Gamma_\nu$. Consequently there is contained in $\Delta \times \tilde{\Delta}$ a proper scalar, an improper vector, a proper tensor of rank 2, etc.

The $n = (2\nu + 1)$ -dimensional group of rotations \mathfrak{d}_n comprises the $(n - 1)$ -dimensional one \mathfrak{d}_{n-1} by subjecting the variables $x_1, \dots, x_{2\nu}$ to an orthogonal transformation and leaving $x_{2\nu+1}$ unchanged. This restriction to a subgroup carries the representation Δ of \mathfrak{d}_n , as here defined, over into the representation Δ of the $(n - 1)$ -dimensional group of rotations which we defined in § 3. The same restriction splits a tensor of rank f in the n -dimensional space into two tensors of rank f and $f - 1$ respectively in the $(n - 1)$ -dimensional space. And thus the decomposition (22) goes over into the decomposition (15).

The matrix C , (17), which satisfied the equations $P'_i = CP_iC^{-1}$ (for $i = 1, 2, \dots, 2\nu$) fulfills the condition

$$CP_nC^{-1} = (-1)^\nu P'_n$$

for $P_n = P_{2\nu+1}$. Hence it can be used here for the same purpose as in § 4 only if ν even. In the opposite case one must replace C by CP_n :

$$C = \begin{vmatrix} 0 & i \\ -i & 0 \end{vmatrix} \times \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \times \begin{vmatrix} 0 & i \\ -i & 0 \end{vmatrix} \times \dots,$$

and one then has $CP_iC^{-1} = -P'_i$ (for all i). Under both circumstances the equation (18) obtains for the C determined in this manner and after an appropriate normalization of the gauge factor in $S(o)$. Here again we have $\tilde{\Delta} \sim \Delta$ and we are able to express explicitly the transformation C which changes covariant spinors into contravariant ones.

6. *Splitting of Δ under restriction to proper rotations.* In the case of odd dimensionality it makes no difference whether one considers the group \mathfrak{d}_n or \mathfrak{d}_n^+ since the reflection commuting with all rotations is an improper rotation. If, however, $n = 2\nu$ is even, restriction to \mathfrak{d}_n^+ effects a splitting of the spin representation Δ into two inequivalent representations Δ^+ and Δ^- of degree $2^{\nu-1}$, and one will have to distinguish between "positive" and "negative" spinors accordingly. This comes about as follows.

Again we form

$$(23) \quad u = \epsilon_{p_1} \dots \epsilon_{p_{2\nu}} \rightarrow U = 1' \times 1' \times \dots \times 1'.$$

We separate the even combinations of signs $(\sigma_1, \dots, \sigma_\nu)$ as characterized by $\sigma_1 \dots \sigma_\nu = +1$ from the odd ones. According to such an arrangement U appears in the form

$$(24) \quad U = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

As a consequence of equations (12) one has for the proper rotations o : $U \rightarrow U^* = U$. As $P_i^* = S P_i S^{-1}$ implies $U^* = S U S^{-1}$ the matrix S commutes with (24) and thus breaks up into an "even" and an "odd" part:

$$S = \begin{bmatrix} S^+ & 0 \\ 0 & S^- \end{bmatrix}.$$

The matrices $S^+(o)$ and $S^-(o)$ in the two representations Δ^+ and Δ^- of degree 2^{n-1} are uniquely determined but for a common sign. Hence the fact that the reflection is associated with the matrix $+1$ in Δ^+ , with the matrix -1 in Δ^- , means an actual inequivalence.

What is the significance of the partition of X into four squares for the corresponding quantities x of the algebra Π or for the tensor sets? (1) We see from the equation $U P_i = -P_i U$ that the even quantities commute with U and that the odd ones anticommute. Even and odd quantities are consequently represented by matrices of the following shape respectively:

$$(25) \quad \begin{bmatrix} \times & \\ & \\ & \times \end{bmatrix},$$

$$(26) \quad \begin{bmatrix} & \times \\ \times & \end{bmatrix}$$

(the squares not marked by a cross are occupied by zeros). (2) The involutorial operation

$$a \rightarrow a^* = aU, \quad A \rightarrow A^* = AU$$

leaves the two front squares in

$$A = \begin{bmatrix} + & - \\ + & - \end{bmatrix}$$

unchanged while it reverses the signs in the two back squares. Let us agree to ascribe the signature $+$ or $-$ to a quantity a according as $a^* = a$ or $a^* = -a$. These quantities then are represented by matrices of the form (27), (28) respectively:

$$(27) \quad \begin{array}{|c|c|} \hline \times & \\ \hline \times & \\ \hline \end{array},$$

$$(28) \quad \begin{array}{|c|c|} \hline & \times \\ \hline & \times \\ \hline \end{array}.$$

Every quantity may be uniquely written as the sum of two quantities of signatures $+$ and $-$. (Besides the operation $a \rightarrow a^*$ one could of course also consider the following one: $a \rightarrow a^\dagger = ua$. But the crossing of both signatures is carried out in a more convenient way by crossing the signature here applied with the division into even and odd quantities. For we have $a^\dagger = a^*$ for even quantities and $a^\dagger = -a^*$ for odd ones.) Thus we finally get this scheme:

$$\begin{array}{cc} \begin{array}{|c|c|} \hline \times & \\ \hline & \\ \hline \end{array} & \begin{array}{|c|c|} \hline & \times \\ \hline & \\ \hline \end{array} \\ \text{even} & \text{odd} \\ + & - \end{array} \quad \begin{array}{cc} \begin{array}{|c|c|} \hline & \\ \hline \times & \\ \hline \end{array} & \begin{array}{|c|c|} \hline & \\ \hline & \times \\ \hline \end{array} \\ \text{odd} & \text{even} \\ + & - : \text{signature.} \end{array}$$

The question as to how our star operation is expressed in terms of tensor sets is answered by the equation:

$$p_1 \cdots p_f \cdot u = (-1)^{\langle f \rangle} \cdot p_{f+1} \cdots p_n,$$

showing that the transition from $a = \{\alpha\}$ to $a^* = \{\alpha^*\}$ is defined by

$$\alpha^*(i'_1 \cdots i'_{n-f}) = (-1)^{\langle f \rangle} \cdot \alpha(i_1 \cdots i_f)$$

(where $i_1 \cdots i_f i'_1 \cdots i'_{n-f}$ is any even permutation). The factor $(-1)^{\langle \nu \rangle} \cdot i$ equals i^ν .

Hence, taking into consideration the splitting of Γ_ν into $\Gamma_\nu^+ + \Gamma_\nu^-$ as explained in § 1, we get the following reductions:

$$(29) \quad \begin{array}{l} \Delta^+ \times \check{\Delta}^+ \sim \Gamma_0 + \Gamma_2 + \cdots \\ \Delta^- \times \check{\Delta}^+ \sim \Gamma_1 + \Gamma_3 + \cdots \end{array} \quad \left| \quad \begin{array}{l} \Delta^+ \times \check{\Delta}^- \sim \Gamma_1 + \Gamma_3 + \cdots \\ \Delta^- \times \check{\Delta}^- \sim \Gamma_0 + \Gamma_2 + \cdots \end{array} \right.$$

Of the two sums in the first column, one breaks off with $\Gamma_{\nu-1}$, the other with Γ_ν^+ , whereas the sums of the second column end with Γ_ν^- and $\Gamma_{\nu-1}$ respectively.

From (16) we obtain by multiplication

$$(-1)^\nu U' = CUC^{-1} \quad \text{or} \quad CU = (-1)^\nu UC.$$

This shows that C is of form (25) or (26) according as ν is even or odd. With C_1, C_2 being the partial matrices of C , we thus have

$$\left. \begin{aligned} \check{S}^+(o) &= C_1 S^+(o) C_1^{-1}, & \check{S}^-(o) &= C_2 S^-(o) C_2^{-1} \\ \check{\Delta}^+ &\sim \Delta^+, & \check{\Delta}^- &\sim \Delta^- \end{aligned} \right\} \quad (\nu \text{ even}),$$

$$\left. \begin{aligned} \check{S}^+(o) &= C_1 S^-(o) C_1^{-1}, & \check{S}^-(o) &= C_2 S^+(o) C_2^{-1} \\ \check{\Delta}^+ &\sim \Delta^-, & \check{\Delta}^- &\sim \Delta^+ \end{aligned} \right\} \quad (\nu \text{ odd}).$$

7. Infinitesimal description.

Even number of dimensions. For the purpose of infinitesimal description it is more convenient to put the quadratic form which is to be left invariant by the orthogonal transformations into the shape

$$(30) \quad x^1 y^1 + x^2 y^2 + \cdots + x^\nu y^\nu.$$

(x^a, y^a being the $n = 2\nu$ variables). Correspondingly one will have to use the following quantities instead of p_a, q_a :

$$\frac{p_a - iq_a}{2} = s_a, \quad \frac{p_a + iq_a}{2} = t_a$$

with the relations

$$\begin{aligned} s_\alpha t_\alpha + t_\alpha s_\alpha &= 1, & s_\alpha t_\beta + t_\beta s_\alpha &= 0 & (\text{for } \beta \neq \alpha), \\ s_\alpha s_\beta + s_\beta s_\alpha &= 0, & t_\alpha t_\beta + t_\beta t_\alpha &= 0 & (\text{for all } \alpha, \beta). \end{aligned}$$

$$s_\alpha \rightarrow S_\alpha = 1' \times \cdots \times 1' \times \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix} \times 1 \times \cdots \times 1,$$

$$t_\alpha \rightarrow T_\alpha = 1' \times \cdots \times 1' \times \begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix} \times 1 \times \cdots \times 1.$$

(The factors written down as matrices stand at the α -th place.)

All infinitesimal rotations are linear combinations of rotations of the following types:

$$(a): \quad dx_\alpha = x_\alpha, \quad dy_\alpha = -y_\alpha;$$

$$(b): \quad dx_\alpha = x_\beta, \quad dy_\beta = -y_\alpha \quad (\alpha < \beta).$$

(The increments not written down are 0. In (b) one is allowed to exchange independently of each other x_α with y_α and x_β with y_β .) Δ represents (a) by the infinitesimal transformation

$$(31) \quad \frac{1}{2} U_\alpha = \frac{1}{2} (1 \times \cdots \times 1 \times 1' \times 1 \times \cdots \times 1)$$

whereas to the infinitesimal rotation (b) corresponds the matrix $S_\alpha T_\beta$. In order to prove this the only thing to be done is to verify the following equations:

$$(a): \quad dX = \frac{1}{2} [U_\alpha, X] = \frac{1}{2} (U_\alpha X - X U_\alpha) = 0$$

for $X = S_\beta$ or T_β ($\beta \neq \alpha$), but $dS_\alpha = S_\alpha$, $dT_\alpha = -T_\alpha$.

$$(b): \quad \delta X = [S_\alpha T_\beta, X] = 0 \quad \text{for all } S \text{ and } T$$

except for $X = S_\beta$ and T_α for which we have:

$$\delta S_\beta = S_\alpha, \quad \delta T_\alpha = -T_\beta.$$

This is readily seen from the expression

$$[S_\alpha T_\beta, X] = S_\alpha(T_\beta X + XT_\beta) - (XS_\alpha + S_\alpha X)T_\beta.$$

In this way we have arrived at Cartan's infinitesimal description of the spin representation.

Nothing essential has to be added in the case of *odd dimensionality*. It is then most convenient to assume the fundamental quadratic form in the shape

$$(x^0)^2 + 2(x^1 y^1 + \dots + x^v y^v).$$

(31) shows that Δ is *double-valued and not single-valued*. For in accordance with this equation the rotation o :

$$x^1 \rightarrow e^{i\phi} x^1, \quad y^1 \rightarrow e^{-i\phi} y^1 \quad (\text{all other variables unchanged})$$

is associated with the operation $S(o)$ multiplying the variable $x_{\sigma_1} \dots x_{\sigma_\nu}$ in the spin space by $e^{i\sigma_1 \phi}$ ($\sigma_\alpha = \pm 1$).

8. *Conditions of reality*. For the *real* orthogonal transformations the question arises whether the conjugate complex representation $\bar{\Delta}: o \rightarrow \bar{S}(o)$ is equivalent to Δ . The P_i being Hermitian matrices, \bar{P}_i equals P_i . Furthermore, the equations:

$$P^*_i = \sum_k o(ki) P_k \quad \text{imply} \quad \bar{P}^*_i = \sum_k o(ki) \bar{P}_k$$

provided the $o(ik)$ are real. This leads at once to the result

$$\bar{S}(o) = \rho(o) \check{S}(o).$$

Hence the Hermitian unit form $\Sigma x_A \bar{x}_A$ in spin space goes over, by means of the substitution S , into ρ fold the unit form. So ρ must be positive and

$$|\det S|^2 = \rho^{2\nu}.$$

But on account of our normalization of S causing $(\det S)^2$ to be $= 1$ we find $\rho = 1$,

$$\bar{S}(o) = \check{S}(o), \quad \bar{\Delta} = \check{\Delta};$$

i. e. the representation Δ of the real orthogonal group is unitary.

When restricting oneself to real variables one must be aware of the possibility that the fundamental quadratic form

$$(32) \quad \sum_{i,k=1}^n a_{ik} x^i x^k$$

may have an *inertial index* t different from 0. This is of particular import for physics as, according to relativity theory, $t = 1$ for the four-dimensional world. One now has to subject the determining p_i of the algebra Π to the equation

$$(p_1 x^1 + \cdots + p_n x^n)^2 = \sum a_{ik} x^i x^k \quad \text{or} \quad \frac{1}{2}(p_i p_k + p_k p_i) = a_{ik}.$$

One will get the new p_i from the old ones by means of the transformation H' if the fundamental form (32) arises from the normal form with $a_{ik} = \delta_{ik}$ by means of the transformation H .

But here again it is convenient to base a more detailed investigation upon the real normal form

$$(33) \quad -(x^1)^2 - \cdots - (x^t)^2 + (x^{t+1})^2 + \cdots + (x^n)^2 = \sum_i \epsilon_i (x^i)^2.$$

(Without any loss of generality we may suppose $2t \leq n$.) In accordance with physics, let us call the first t variables x^i the temporal, the last $n - t$ the spatial coördinates. The subject of our consideration is the group \mathfrak{h}_n of Lorentz transformations; that is, of all real linear transformations o carrying the fundamental form (33) into itself.†

P_{t+1}, \dots, P_n keep their previous significance, while P_1, \dots, P_t assume the factor $i = \sqrt{-1}$. We thus have

$$\bar{P}_i = -P'_i \quad \text{for} \quad (i = 1, \dots, t); \quad \bar{P}_i = P'_i \quad \text{for} \quad (i = t + 1, \dots, n).$$

The Hermitian conjugate \bar{A}' of a matrix A may be denoted by \bar{A} . The \bar{P}_i as well as the P'_i satisfy the fundamental rules of commutation. Both sets of matrices must be changed one into the other by means of a certain transformation B . It is easy enough to write down B explicitly:

$$(34) \quad B = i^{t-\langle t \rangle} \cdot P_1 \cdots P_t.$$

To be exact, we have

† To be quite definite: the variables x^i are subjected to the Lorentz-transformation $o: x^i \rightarrow \sum_k o(ik) x^k$. The p_i (or P_i) then undergo the contragredient transformation; but in raising the index by means of $p^i = \epsilon_i p_i$ one may introduce quantities p^i transforming cogrediently with the variables x^i .

$$(35) \quad P'_i = B\bar{P}_i B^{-1} \quad \text{or} \quad -P'_i = B\bar{P}_i B^{-1}$$

according as t is even or odd. The factor $i^{t-\langle t \rangle}$ has been added in order to make B Hermitian: $\bar{B} = B$. The transposed matrix B' coincides with B but for the sign, namely $B' = (-1)^{\langle t \rangle} B$. In the case of an even n the matrix B is of form (25) or (26) according as t is even or odd. All these properties could be fairly easily derived from general considerations; it is not worth the trouble, however, as one may read them at once from the explicit expression (34).

One obtains from (35) the relation

$$(36) \quad B\bar{S}(o)B^{-1} = \rho(o)\bar{S}(o)$$

or after multiplication by $S'(o)$ on the left:

$$S'B\bar{S} = \rho B;$$

the Hermitian form B goes over, by means of the transformation \bar{S} , into the multiple ρ of itself. In consequence ρ is real and one infers, in the same manner as in the definite case, the equation

$$\rho(o) = \pm 1.$$

As to its dependence on o , $\rho(o)$ satisfies the condition

$$\rho(o'o) = \rho(o')\rho(o).$$

A new consideration, however, is required for determining this sign ρ . In a Lorentz transformation $\|o(ik)\|$ the temporal minor of the whole determinant:

$$(37) \quad \Omega = \begin{vmatrix} o(11), & \dots, & o(1t) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ o(t1), & \dots, & o(tt) \end{vmatrix} \quad \begin{array}{l} \text{is either } \geq 1 \\ \text{or } \leq -1. \end{array}$$

We shall put $\sigma_-(o) = +1$ or -1 according as the first or the second case prevails, and call $\sigma_-(o)$ the *temporal signature*; it is a character, i. e.

$$\sigma_-(o'o) = \sigma_-(o') \cdot \sigma_-(o).$$

We need not trouble to prove this here directly because we shall see in the course of our further investigations that the $\rho(o)$ in (36) coincides with $\sigma_-(o)$. In the same manner one may introduce a *spatial signature* $\sigma_+(o)$ by means of the spatial minor of the matrix $\|o(ik)\|$. The latter, though, is $= \sigma(o) \cdot \Omega$;

hence the character $\sigma(o)$ distinguishing the proper and improper transformations equals $\sigma_+ \sigma_-$. Of the Lorentz transformations having $\sigma_- = -1$ one may say that they reverse the sense of time whereas those having $\sigma_+ = -1$ reverse the spatial sense. The group of Lorentz transformations falls apart into four pieces not connected with each other and distinguished from each other by the values of the two signatures σ_- and σ_+ .

To prove (37) let us introduce the two vectors

$$v_i' = \{o(i1), \dots, o(it)\}, \quad v_i'' = \{o(i, t+1), \dots, o(in)\}$$

in the realms of the temporal and spatial coördinates respectively. The scalar product $(a' \cdot b')$ in these two partial spaces has its usual significance $a'_1 b'_1 + \dots + a'_t b'_t$. The relations characteristic for the Lorentz transformation then read:

$$(v_i' v_k') = \delta_{ik} + (v_i'' v_k'') \quad (i, k = 1, 2, \dots, t).$$

From these we derive

$$\begin{aligned} \begin{vmatrix} (v_1' v_1') & \dots & (v_1' v_t') \\ \vdots & & \vdots \\ (v_t' v_1') & \dots & (v_t' v_t') \end{vmatrix} &= \begin{vmatrix} 1 + (v_1'' v_1'') & (v_1'' v_2'') & \dots & (v_1'' v_t'') \\ \vdots & \vdots & \ddots & \vdots \\ (v_t'' v_1'') & (v_t'' v_2'') & \dots & (1 + v_t'' v_t'') \end{vmatrix} \\ &= 1 + (1/1!) \sum_{i=1}^t (v_i'' v_i'') + (1/2!) \sum_{i,k=1}^t \begin{vmatrix} (v_i'' v_i'') & (v_i'' v_k'') \\ (v_k'' v_i'') & (v_k'' v_k'') \end{vmatrix} + \dots \end{aligned}$$

All terms on the right side are ≥ 0 ; hence the whole determinant on the left is ≥ 1 . This determinant however is the square of Ω .

The fact that the sign ρ in (36) equals σ_- is proved in the following manner. In accordance with

$$P^*_i = \sum_{k=1}^n o(ki) P_k$$

we find

$$(38) \quad P^*_1 \dots P^*_t = \begin{vmatrix} o(11) & \dots & o(1t) \\ \vdots & & \vdots \\ o(t1) & \dots & o(tt) \end{vmatrix} \cdot P_1 \dots P_t + \dots$$

But a product like $P_{i_1} \dots P_{i_t} \cdot P_1 \dots P_t$ where $i_1 \dots i_t$ are different indices always has the trace 0 except if $i_1 \dots i_t$ is a permutation of $1 \dots t$; whereas

$$\text{tr}(P_1 \dots P_t \cdot P_1 \dots P_t) = (-1)^{\langle t \rangle} \text{tr}(P_1^2 \dots P_t^2) = (-1)^{t - \langle t \rangle} \cdot 2^v.$$

Hence on multiplying equation (38) by $P_1 \dots P_t$ to the right and forming the trace, one is led to this value of the determinant Ω :

$$2^v \Omega = (-1)^{t - \langle t \rangle} \text{tr}(P^*_1 \dots P^*_t \cdot P_1 \dots P_t).$$

Using the definitions of S : $P^*_i = SP_i S^{-1}$, and of B , one readily obtains:

$$2^v \Omega = \text{tr}(SBS^{-1} \cdot B) = \text{tr}(B \cdot SBS^{-1}).$$

According to (36)

$$S^{-1} = {}_\rho B'^{-1} \tilde{S} B' = {}_\rho B^{-1} \tilde{S} B.$$

Replacement of B' by B is allowed as B' coincides with B but for a numerical factor. So one finally gets, with $T = BS = \|t_{JK}\|$:

$$2^v \Omega = \rho \cdot \text{tr}(BS \tilde{S} B) = \rho \cdot \text{tr}(BS \cdot \tilde{S} \bar{B}) = \rho \cdot \text{tr}(T \cdot \bar{T}) = \rho \cdot \sum_{J,K} |t_{JK}|^2,$$

and this equation shows ρ to have the sign of Ω .

Any representation $\Gamma: o \rightarrow G(o)$ of the Lorentz group gives rise to another one $\sigma_- \Gamma: o \rightarrow \sigma_-(o)G(o)$. Equation (36) or

$$\tilde{S}(o) = \sigma_-(o) B^{-1} \tilde{S}(o) B$$

then proves the equivalence:

$$(39) \quad \bar{\Delta} \sim \sigma_- \check{\Delta}.$$

The transformation B changes the conjugate of a covariant spinor ψ into a contravariant spinor $\phi: \phi' = B\bar{\psi}$ (in so far as we confine ourselves to Lorentz's transformations of temporal signature $\sigma_- = 1$). (39) yields, on account of (15), (22), the decompositions

$$(40) \quad \Delta \times \bar{\Delta} \sim \left\{ \begin{array}{l} \sigma_- \Gamma_0 + \sigma_- \Gamma_1 + \cdots + \sigma_- \Gamma_{v-1} + \\ \sigma_+ \Gamma_0 + \sigma_+ \Gamma_1 + \cdots + \sigma_+ \Gamma_{v-1} \end{array} \right\} + (\sigma_- \Gamma_v \sim \sigma_+ \Gamma_v) [n = 2v];$$

$$\Delta \times \bar{\Delta} \sim \sigma_- \Gamma_0 + \sigma_+ \Gamma_1 + \sigma_- \Gamma_2 + \cdots \quad [n = 2v + 1].$$

The latter series breaks off with $\sigma_- \Gamma_v$ or $\sigma_+ \Gamma_v$.

In the case $n = 2v$ we have the splitting of Δ into Δ^+ and Δ^- , when restricting ourselves to the group \mathfrak{d}_n^+ of proper Lorentz transformations [$\sigma(o) = 1$]. This restriction wipes out the difference between the two signatures σ_- and σ_+ . As we mentioned before, B is of form (25) or (26) according as t is even or odd. Hence one has

$$\begin{array}{ll} \text{for even } t: & \bar{\Delta}^+ \sim \sigma_- \check{\Delta}^+, \quad \bar{\Delta}^- \sim \sigma_- \check{\Delta}^-; \\ \text{for odd } t: & \bar{\Delta}^+ \sim \sigma_- \check{\Delta}^-, \quad \bar{\Delta}^- \sim \sigma_- \check{\Delta}^+. \end{array}$$

9. *Irreducibility.* Irreducibility of Γ_f is granted *a fortiori* if one is able to prove that there does not exist any homogeneous linear relation with constant coefficients (independent of o) among the minors of order f of the matrix of an arbitrary rotation $\|o(ik)\|$. This can be shown without using

any other rotations than permutations of the coördinate axes combined with changes of signs. For let us assume that we have such a non-trivial relation R in which a definite minor $A(i_1 \cdots i_f)$ occurs with a coefficient different from 0. By suitable exchange we can place this minor in the left upper corner of the matrix. We will now take into account the changes of signs only:

$$\|o(ik)\| = \begin{vmatrix} \pm 1 & & & \\ & \pm 1 & & \\ & & \ddots & \\ & & & \pm 1 \end{vmatrix}$$

the matrices of which have only their chief minors $A(i_1 \cdots i_f)$ different from 0. The linear relation R will contain, apart from $A(1\ 2 \cdots f)$, at least one more term $A(1'\ 2' \cdots f')$ with a coefficient different from zero. At least one of the indices $1'\ 2' \cdots f'$, let us say l , is different from $1, 2, \cdots, f$. By changing the sign of the one variable x_l , the relation R is carried over into a new one R' in which $A(1\ 2 \cdots f)$ occurs with the same, $A(1'\ 2' \cdots f')$ however with the opposite coefficient. Hence the sum $\frac{1}{2}(R + R')$ certainly is shorter than R , that is, contains less terms than R ; but $A(1\ 2 \cdots f)$ occurs in it with the same coefficient different from 0 as before. The procedure of shortening may be continued until the presupposed linear relation $R = 0$ leads to the impossible equation $A(1\ 2 \cdots f) = 0$.

These considerations were based upon the *complete* group \mathfrak{d}_n . If one allows proper rotations only, \mathfrak{d}_n^+ , one may have to combine the permutation in the first step with a change of sign of one variable. The second step can be performed in the same manner provided $2f < n$, for then one may choose l as above: as one of the indices $1', 2', \cdots, f'$ different from $1, 2, \cdots, f$, furthermore choose m as an index that does not occur in the row $1, 2, \cdots, f, 1', 2', \cdots, f'$, and then change the signs of both variables x_l and x_m simultaneously. Even when $n = 2v$, $f = v$ the procedure of shortening will work as long as the relation R still contains a term $A(1'\ 2' \cdots v')$ the indices of which are not just the complement $v + 1, \cdots, n$ of $1, \cdots, v$. Thus one will be led in this case finally to a relation of the form:

$$(41) \quad c A(1, 2, \cdots, v) + c' A(v + 1, \cdots, n) = 0.$$

Such a relation obtains indeed:

$$A(v + 1, \cdots, n) = A(1, 2, \cdots, v)$$

but there exists of course no other one of the type (41). From this we learn not only that the two representations Γ_ν^+ and Γ_ν^- are irreducible, but at the same time that they are *inequivalent*; for it proves that there does not hold any linear relation with fixed coefficients between the components of the two matrices associated with the same arbitrary rotation σ in these representations. For the components of these two matrices are

$$\frac{1}{2} [B(\begin{smallmatrix} i_1 \dots i_\nu \\ k_1 \dots k_\nu \end{smallmatrix}) \pm i^\nu B(\begin{smallmatrix} i'_1 \dots i'_\nu \\ k_1 \dots k_\nu \end{smallmatrix})]$$

with

$$B(\begin{smallmatrix} i_1 \dots i_\nu \\ k_1 \dots k_\nu \end{smallmatrix}) = \frac{1}{2} [A(\begin{smallmatrix} i_1 \dots i_\nu \\ k_1 \dots k_\nu \end{smallmatrix}) + A(\begin{smallmatrix} i'_1 \dots i'_\nu \\ k'_1 \dots k'_\nu \end{smallmatrix})].$$

$i_1 \dots i_\nu, i'_1 \dots i'_\nu$ and $k_1 \dots k_\nu, k'_1 \dots k'_\nu$ are even permutations of the figures $1, 2, \dots, n$. The reasoning above shows that there exists no universal linear relation between the quantities $B(\begin{smallmatrix} i_1 \dots i_\nu \\ k_1 \dots k_\nu \end{smallmatrix})$.

The *inequivalence* of two such Γ_f the ranks f of which do not give the sum n , is granted by their having different degrees.

This whole argument was based upon the *complex* orthogonal group. But nothing is to be modified when one confines oneself to the *real* orthogonal transformations. Furthermore one sees, by formulating the result in an infinitesimal manner, that it cannot be effected by the inertial index. The infinitesimal transformation

$$(42) \quad dx_i = x_k, \quad dx_k = -x_i \quad (i \neq k)$$

(all other increments being 0; this transformation engenders the permutation $x_i \rightarrow x_k, x_k \rightarrow -x_i$ as well as the change of sign $x_i \rightarrow -x_i, x_k \rightarrow -x_k$) has to be replaced, if the fundamental quadratic form contains terms with the minus sign, for couples (x_i, x_k) consisting of a temporal and a spatial variable by

$$dx_i = x_k, \quad dx_k = x_i$$

while it has to be kept unchanged for couples of variables (x_i, x_k) both temporal or both spatial. The statement of irreducibility under all transformations (42) in the definite case is identical with the statement of irreducibility under the transformations replacing them in the indefinite case; one only needs to replace the temporal variables x_k by $\sqrt{-1} \cdot x_k$.

The product $\Gamma \times \check{\Gamma}$ of a representation Γ with its contragredient $\check{\Gamma}$ contains the identity Γ_0 at least μ times when Γ reduces into μ parts. If we are allowed to make use of the general and elementary theorem that the irreducible

parts of a representation are uniquely determined † (in the sense of equivalence and except for their arrangement), then the formulae (15), (22), (29) show at once the irreducibility of Δ or Δ^+ and Δ^- respectively and the inequivalence of the latter. Another direct proof runs as follows:

Take the full group \mathfrak{d}_n in the even case $n = 2\nu$. Using the fundamental quadratic form in the shape (30), let us consider the “diagonal” infinitesimal rotations

$$(43) \quad dx_a = i\phi_a x_a, \quad dy_a = -i\phi_a y_a \quad (\alpha = 1; \dots, \nu)$$

(ϕ_a independent parameters). It is associated in Δ with the diagonal transformation

$$dx_{\sigma_1 \dots \sigma_\nu} = (i/2)(\sigma_1 \phi_1 + \dots + \sigma_\nu \phi_\nu) x_{\sigma_1 \dots \sigma_\nu} \quad (\sigma_a = \pm).$$

Given a partial space P' of the total spin space P , different from 0 and invariant under Δ , one chooses a non-vanishing vector z :

$$z = \sum_A z_A e_A = \{z_A\} \quad [A = (\sigma_1, \dots, \sigma_\nu)]$$

occurring in P' . By performing the substitution (43) repeatedly one is able to isolate each term $z_A e_A$, as these parts are of different “weights” $(i/2)(\sigma_1 \phi_1 + \dots + \sigma_\nu \phi_\nu)$. Therefore at least one of the fundamental vectors e_A occurs in P' . But $e_A = e_{\sigma_1 \dots \sigma_\nu}$ goes over into any other fundamental vector $e_{\tau_1 \dots \tau_\nu}$ by exchanging $x_a \rightarrow y_a$, $y_a \rightarrow x_a$ those couples (x_a, y_a) for which the signs σ_a and τ_a do not coincide. P' is therefore identical with the total P .—Irreducibility of Δ for odd $n = 2\nu + 1$ is an immediate consequence of the irreducibility for even n , we just proved; one has to restrict oneself merely to the subgroup \mathfrak{d}_{n-1} within \mathfrak{d}_n , $n = 2\nu + 1$. One sees in the same manner that the two parts Δ^+ , Δ^- are irreducible and inequivalent for the group \mathfrak{d}_n^+ , $n = 2\nu$.

10. *Dirac's theory.* Let us suppose we are dealing with a *spinor field* $\psi^A(x^1 \dots x^n)$ in an n -dimensional “world” with the fundamental metric form (33). The most essential feature of *Dirac's theory* is that one should be able to form a *vector* by linear combination of the products $\bar{\psi}^A \psi^B$. If n is even, one sees from equation (40) that exactly *one* such vector s_i exists—that behaves like a vector at least for all Lorentz transformations not reversing the sense of time; and *one* such vector for all Lorentz transformations not reversing the spatial sense. In the case n odd, one vector of the second, and

† Compare e.g. Weyl, *Theory of Groups and Quantum Mechanics* (London, 1931), p. 136.

no vector of the first kind exists. Only the first type can be used when one believes in the equivalence of right and left, but is prepared to abandon the equivalence of past and future. n has then to be even and the vector is

$$s_i = \bar{\psi} B P_i \psi.$$

From this vector one can derive the scalar field:

$$(44) \quad \sum_i \bar{\psi} B P^i (\partial \psi / \partial x^i) \quad (P^i = \epsilon_i P_i).$$

One needs a scalar that arises from linear combination of the products $\bar{\psi}^A \cdot \partial \psi^B / \partial x^i$ in Dirac's theory as the main part of the *action quantity* which accounts for the fundamental features of the whole quantum theory. There is no ambiguity: for $(\Delta \times \bar{\Delta}) \times \check{\Gamma}_1$ contains the identity Γ_0 or rather the representation $\sigma \Gamma_0$ just once if decomposed into its irreducible parts. That is shown by equation (40) when one takes into account the fundamental lemma of the theory of representations asserting that the product $\Gamma \times \check{\Gamma}_1$ contains the identity Γ_0 once, or not at all, according as the two irreducible representations Γ, Γ_1 of the same group are equivalent or not. Dirac's quantity of action contains, apart from (44), a second term which is a linear combination of the undifferentiated products $\bar{\psi}^A \psi^B$; it is multiplied by the mass, and accounts for the inertia of matter. There exists just one such scalar, namely $\bar{\psi} B \psi$, in the case of an even as well as an odd n .

Furthermore one may consider as essential the fact that the time component of the electric current is positive-definite in Dirac's theory, namely proportional to the "probability density" $\sum_A \bar{\psi}^A \psi^A$; this grants the atomistic structure of electric charge. If the fundamental form (33) is of inertial index t , this property however is not possessed by the vector contained in $\Delta \times \bar{\Delta}$ but by the tensor of rank t with the components

$$s_{i_1 \dots i_t} = \bar{\psi} B P_{i_1} \dots P_{i_t} \psi \quad (i_1, \dots, i_t \text{ different}),$$

the "temporal" component, $s_{12 \dots t}$, of which is $= \bar{\psi} \psi$ (but for a numerical factor). It seems to be required by the scheme of Maxwell's equations that electric current should be a vector; this requirement, together with the postulate of the atomic structure of electricity, compels us to assume the inertial index t to be $= 1$.

11. Appendix. Automorphisms of the complete matrix algebra. A one-

to-one correspondence $X \rightleftharpoons X^*$ of the ring of all n -rowed matrices upon itself is isomorphic when satisfying the conditions

$$(X + Y)^* = X^* + Y^*, \quad (\lambda X)^* = \lambda \cdot X^*, \quad (XY)^* = X^*Y^*$$

(λ an arbitrary number). The only such automorphism is "similarity":

$$X^* = AXA^{-1},$$

A being a fixed non-singular matrix.

Proof. The equation $GX = \gamma X$ has a solution $X \neq 0$ only if γ is an eigen-value of the matrix G ; for the columns of the matrix X must be eigenvectors belonging to the eigen-value γ . The eigen-values of G thus are characterized in a manner invariant with respect to the given automorphism. Consequently G^* has the same eigen-values as G . Thus we are led to proceed as follows. Let us choose n fixed different numbers $\gamma_1, \dots, \gamma_n$ and with them form the diagonal matrix

$$G = \begin{vmatrix} \gamma_1 & & \\ & \ddots & \\ & & \gamma_n \end{vmatrix}.$$

As G^* has the same eigen-values as G , a non-singular matrix A can be determined such that $G^* = AGA^{-1}$. Let us replace every X^* by $X^{**} = A^{-1}X^*A$ and now consider the automorphism $X \rightarrow X^{**}$ that leaves G unchanged. The matrix E_{ik} containing an element different from 0, namely 1, only at the crossing point of the i -th row with the k -th column is determined by the properties

$$GE_{ik} = \gamma_i E_{ik}, \quad E_{ik}G = \gamma_k E_{ik}$$

except for a numerical factor. Hence we have

$$(45) \quad E_{ik} \rightarrow E_{ik}^{**} = \alpha_{ik} E_{ik}.$$

The equation $E^2_{ii} = E_{ii}$ furnishes $\alpha^2_{ii} = \alpha_{ii}$, $\alpha_{ii} = 1$. After putting $\alpha_{i1} = \alpha_i$, $\alpha_{1k} = \beta_k$, the relation

$$E_{ik} = E_{i1}E_{1k}$$

leads to $\alpha_{ik} = \alpha_i \beta_k$. On account of $\alpha_{ii} = 1$ one therefore has $\beta_i = 1/\alpha_i$ and $\alpha_{ik} = \alpha_i/\alpha_k$. Hence in accordance with (45) an arbitrary matrix $X = \|x_{ik}\|$ and its image $X^{**} = \|x^{**}_{ik}\|$ are linked by the relation

$$x_{ik}^{**} = \alpha_i x_{ik} / \alpha_k \quad \text{or} \quad X^{**} = A_0 X A_0^{-1}$$

where A_0 is the diagonal matrix with the terms $\alpha_1, \dots, \alpha_n$.

This demonstration furnishes a method for constructing a spinor from a given tensor set g . The method will be used preferably in the case where g consists of only one tensor of definite rank. Our representation of degree 2^r of the algebra Π associates with g a matrix G . Let us assume that G has the (simple) eigen-value γ and let ψ be the corresponding eigen-vector in spin space: $G\psi = \gamma \cdot \psi$. The rotation o carries g into a set $g(o)$ represented by the matrix $G(o)$. γ is a (simple) eigen-value of $G(o)$ as well as of G , and the solution $\psi(o)$ of the equation

$$G(o)\psi(o) = \gamma \cdot \psi(o)$$

arises from ψ by the transformation $S(o)$ corresponding to o in the spin representation.

THE INSTITUTE FOR ADVANCED STUDY,
PRINCETON, NEW JERSEY.

ON THE THEORY OF APPORTIONMENT.

By WILLIAM R. THOMPSON.

1. If in an accepted sense, P is the probability that one method of treatment, T_1 , is better than a rival, T_2 , we may develop a system of apportionment such that the proportionate use of T_1 is $f_{(P)}$, a monotone increasing function, rather than make no discrimination at all up to a certain point and then finally entirely reject one or the other. The only paper* which has so far appeared in his field, as far as I am aware, is one by myself in a recent issue of *Biometrika*. In this paper I have considered the case of choice between two such rival treatments,† and for symmetry suggested that $f_{(Q)} \equiv 1 - f_{(P)}$, where $Q = 1 - P$. Then the risk of assignment to T_1 when it is not the better is $Q \cdot f_{(P)}$, while the corresponding risk for T_2 is $P \cdot f_{(Q)}$. Accordingly, I suggested further that we set $f_{(P)} = P$, which is a necessary and sufficient condition that these two risks be equal. Their sum, the total risk, is then $2PQ$.

A special case was considered wherein the result of use of T_i at any given trial is either *success* or *failure*, the probability of failure being an unknown, p_i , *a priori* (independently for $i = 1, \dots, k$) equally likely to lie in either of any two equal intervals in the possible range, $(0, 1)$. It is further assumed that for a given T_i we have an experience of exactly n_i independent trials, the number of *successes* being s_i and of *failures* being $r_i \equiv n_i - s_i$; and the probability of obtaining such a sample is

$$\binom{n_i}{r_i} \cdot p_i^{r_i} \cdot q_i^{s_i} \text{ where } q_i = 1 - p_i.$$

Restricting consideration to the case, $k = 2$, dropping the subscript *one* and using a prime instead of subscript *two*, then it was shown that

$$(1) \quad P = \psi_{(r,s,r',s')} \equiv \frac{\sum_{\alpha=0}^{r'} \binom{r+r'-\alpha}{r} \cdot \binom{s+s'+1+\alpha}{s}}{\binom{n+n'+2}{n+1}}.$$

Now, it is well known that the probability, \bar{P} , that by drawing at random

* W. R. Thompson, *Biometrika*, vol. 25 (1933), pp. 285-294.

† By *treatment* we imply a special mode of dealing with *individuals* of a given class of things.

without replacements from a mixture of W white and B black balls we shall encounter w white before b black is given by

$$(2) \quad \bar{P} = \frac{\sum_{\alpha=0}^h \binom{W}{w+\alpha} \cdot \binom{B}{b-1-\alpha}}{\binom{W+B}{w+b-1}},$$

where $h = \text{Min}(b-1, W-w)$. The object of the present paper is *first*, to show exactly how ψ may be expressed in the form of (2) and thus make possible the use of a machine based on this principle in the apportionment, and thereby avoid an enormous amount of calculation where tables are not available; and *second*, to develop a complete statement of the group, G , of substitutions of the arguments of $\psi_{(a_1, a_2, a_3, a_4)}$ which leave ψ invariant, and also those of the set, A , which change the value to $1 - \psi_{(a_1, a_2, a_3, a_4)}$. The application of these substitutions to give a convenient form for calculation* of ψ or for other purposes is obvious. On this account the ψ -function is a convenient form for expression† of the incomplete hypergeometric series, as in the case of two problems considered by Pearson,‡ where for certain original variables which we may denote by a, b, c , and d we may express § a required probability by $\psi_{(a, b, c, d-1)}$.

2. We begin by considering the function, $\bar{N}_{(r, s, r', s')}$ of four rational integers ≥ 0 , defined by

$$(3) \quad \bar{N}_{(r, s, r', s')} \equiv \sum_{\alpha=0}^{a \leq s, r'} \binom{r+r'+1}{r+1+\alpha} \cdot \binom{s+s'+1}{s-\alpha},$$

and extend this definition to include

$$(4) \quad \begin{aligned} \bar{N}_{(r, s, -1, s')} &\equiv 0 \equiv \bar{N}_{(r, -1, r', s')}, \quad \text{and} \\ \bar{N}_{(r, s, r', -1)} &\equiv \binom{r+s+r'+1}{r'} \equiv \bar{N}_{(-1, r', s, r)}. \end{aligned}$$

Now, in the previous paper,¶ I have defined an N -function identical with \bar{N} for the arguments in (4) and otherwise equal to the numerator of the right member of (1). Obviously,

* B. H. Camp, *Biometrika*, vol. 17 (1925), pp. 61-67.

† W. R. Thompson, *loc. cit.*

‡ Karl Pearson, *Philosophical Magazine*, Series 6, vol. 13 (1907), pp. 365-378; *Biometrika*, vol. 20A (1928), pp. 149-174.

§ W. R. Thompson, *loc. cit.*

¶ W. R. Thompson, *loc. cit.*

$$\bar{N}_{(r,s,r',s')} \equiv \bar{N}_{(s',r',s,r)} \equiv \binom{n+n'+2}{n+1} - \bar{N}_{(s,r,s',r')},$$

as has been proved for the N -function,* and

$$(5) \quad \bar{N}_{(r,s,0,s')} \equiv N_{(r,s,0,s')};$$

and we may verify readily by (3) that *in general*

$$(6) \quad \bar{N}_{(r,s,r',s')} \equiv \bar{N}_{(r,s-1,r',s')} + \bar{N}_{(r,s,r',s'-1)},$$

which relation was shown in my first paper to hold for the N -function also. Accordingly, by complete induction we may demonstrate that

$$(7) \quad \bar{N}_{(r,s,r',s')} \equiv N_{(r,s,r',s')},$$

and therefore

$$(8) \quad \psi_{(r,s,r',s')} \equiv \frac{\sum_{\alpha=0}^{a \leq s, r'} \binom{r+r'+1}{r+1+\alpha} \cdot \binom{s+s'+1}{s-\alpha}}{\binom{r+s+r'+s'+2}{r+s+1}};$$

By a simple rearrangement of factors after expressing the binomial coefficients in (8) by factorial numbers we may obtain

$$(9) \quad \psi_{(r,s,r',s')} \equiv \frac{\sum_{\alpha=0}^{a \leq s, r'} \binom{r+s+1}{r+1+\alpha} \cdot \binom{r'+s'+1}{r'-\alpha}}{\binom{r+r'+s+s'+2}{r+r'+1}}$$

which is the equivalent of the expression in (2) if we set $W = r + s + 1$, $B = r' + s' + 1$, $w = r + 1$ and $b = r' + 1$, which is the required relation. Furthermore, (8) and (9) give

$$(10) \quad \psi_{(r,s,r',s')} \equiv \psi_{(r,r',s,s')},$$

i. e., $\psi_{(a_1, a_2, a_3, a_4)}$ is invariant under the substitution (2, 3), which therefore belongs to the group, G . Now, by the identities of (10) and (23) of the previous paper,† we have obviously established that (1, 4)(2, 3) is also in G , and that (1, 2)(3, 4) changes ψ to $1 - \psi$ and is therefore an element of the set A . On the other hand if $a_1 = 3$, $a_2 = 2$, $a_3 = 1$, and $a_4 = 0$, the substitution (1, 3) brings a change in value of ψ from 9/14 to 13/14, and therefore (1, 3) belongs neither to G nor A . Now, if the four arguments are all different they may be arranged in 24 different ways; whence, if m is the

* W. R. Thompson, *loc. cit.*

† W. R. Thompson, *loc. cit.*

number of different substitutions in the group, G , then $24/4 \geq m \equiv 0 \pmod{4}$. Accordingly, we have established the fact that the complete group leaving ψ invariant is generated by the two transpositions, $(2, 3)$, and $(1, 4)$; i. e.,

$$(11) \quad G = [(2, 3), (1, 4)].$$

Moreover, the set of substitutions, A , changing ψ to $1 - \psi$ may be represented in the form,

$$(12) \quad A = \{g \cdot (1, 2)(3, 4)\}$$

where g is an element of G .

By the aid of (11) and (12) we may prove and state in simple form certain relations,* and prior to any use of the ψ -function obtain the most convenient arrangement for the work; and in tabulations only 3 values need be listed for each combination of the four arguments without loss of completeness, namely $\psi_{(a,b,c,d)}$, $\psi_{(a,b,d,c)}$, and $\psi_{(a,c,d,b)}$. We may readily verify also that if two of these arguments are equal then two of the three values are sufficient, if three of the arguments are equal or there are two pairs of equal arguments then one value is enough, and if $a = b = c = d$ then none is needed in order to evaluate ψ in a simple manner by means of (11) and (12). By use of the N -function as previously suggested † instead of ψ intabulation in a systematic process with increasing arguments we may list only values of this reduced form of table; e. g., $a \geq d \geq c \geq b > 0$ with the relations given in (6) and (7) and

$$(13) \quad N_{(r,s,r',0)} \equiv \binom{r+s+r'+2}{r'+1} - \binom{r+r'+1}{r}$$

and $N_{(r,0,r',s')} \equiv \binom{r+r'+1}{r}.$

* W. R. Thompson, *loc. cit.*

† Thus we may obtain readily, the relation,

$$\psi_{(r,s,r',s')} \equiv \psi_{(r-1,s,r',s')} - \frac{(s+s'+1) \cdot \binom{r+r'}{r} \cdot \binom{s+s'}{s}}{(r+s+1) \binom{r+s+r'+s'+2}{r+s+1}},$$

and simply from limit relations previously established,

$$\begin{aligned} I_{p(r+1,s+1)} &\equiv I_{p(r,s+1)} - \binom{n}{s} p^r \cdot q^{s+1} \\ &\equiv I_{p(r+1,s)} + \binom{n}{r} p^{r+1} \cdot q^s \end{aligned}$$

where $q = 1 - p$, and $I_{x(u,v)} \equiv \frac{B_x(u,v)}{B_1(u,v)}.$

3. For my own purposes I constructed a rough machine based on the probability relation (9) as follows:

I took the cover of a square cardboard box, which I cut and bent along the diagonal forming a box having the shape of an isosceles triangle with 45° base angles. In this I placed $n + n' + 2$ balls as used in bearings. Of these $n' + 1$ had been made dull by a copper sulphate bath. I shall call these *black* and the others *white*. I then shuffled these balls in the box, and at random allowed them all * to line up along the long side or hypotenuse of the box. This alignment I regarded as a draft proceeding from left to right. Here the advantage of a prior arrangement of the arguments of ψ so as to make the number of balls to be scanned as small as possible is apparent. The critical condition was to encounter $r + 1$ *white* before $r' + 1$ *black* balls.

I supposed now that I was considering a case of the sort where I have to assign individuals to one of two methods of treatment,† T_1 and T_2 , in proportion based on the ψ -function of the accumulated evidence in the conventional r, s, r', s' form. I then gave certain values to p_1 and p_2 to govern the chance of *failure* when T_1 and T_2 were tried, respectively; but otherwise acted as if p_1 and p_2 were unknown. Starting with no experience, then $r = s = r' = s' = 0$, I placed $r + s + 1 = 1$ white and $r' + s' + 1 = 1$ black ball in the box, and shuffled. After alignment then T_1 was chosen if the white ball was at the left and otherwise T_2 was chosen. The treatment chosen, T_i , was *tried* by the corresponding probability, p_i , and the result recorded in new values of r, s, r', s' ; i. e., if T_2 were tried with *success* these new values then were 0, 0, 0, 1; if with *failure* then they would have been 0, 0, 1, 0. Similar remarks hold if T_1 were chosen. I then added a ball, white if T_1 had been tried and otherwise black. These three balls were now shuffled and aligned at random. As before, if the critical condition of encountering $r + 1$ white before $r' + 1$ black balls were met then the treatment, T_1 was used at this turn, and otherwise T_2 . The result of the treatment indicated was noted and new values of r, s, r', s' obtained, and another ball added to the box according to the criterion described for the last turn, and so on until a given number of trials had been made.

In the accompanying table values of p_1 and p_2 used in such experiments are given together with the final results—the total number of trials, $n + n'$;

* As a matter of fact it is not necessary that all the balls be lined up. The object is simply to quickly establish a random draft order.

† By *treatment* we imply a special mode of dealing with *individuals* of a given class of things.

the number of these wherein the *conventionally* worse method (T_1) was used, n ; and the number of *failures*, r , among these n trials.

To make the table quite clear, take the numbers in the second row. Here we have the record of four parallel experiments wherein T_2 was governed by a condition such that failure might be expected about half the time and T_1 to fail always. The total number of trials, $n + n' = 40$, and the number of these systematically allotted to T_1 was $n = 5, 9, 7$, and 5 in the respective experiments, and r , of course, had the same values here. The relatively small value of these even in so small a total number of trials, indicates strikingly the rapidity with which this systematic apportionment between the rival treatments, T_1 and T_2 , tends to favor the better, even though prior knowledge as to the fact that T_2 is the better is disregarded.

Although the machine used is extremely crude, all the results obtained were extremely favorable. A more carefully constructed machine along the same lines might give even better results. I have conducted a few additional experiments with this simple box, in which I have deliberately arranged an unfavorable start. I was greatly pleased to note the rapidity with which the machine brought about a reversal of favor to the better method, T_2 , as the experiments proceeded.

4. The system of apportionment which we have examined admits a simple extension to the general case of k rival *treatments*, (T_i). As defined in § 1, we let p_i represent the unknown probability of failure by treatment T_i , and our experience with this treatment to consist of r_i failures and s_i successes, where $i = 1, \dots, k$. Now, if we place $r_i + s_i + 1$ balls of a kind, C_i ; for $i = 1, \dots, k$; in our box, shuffle and draw as before, then we note that the probability of drawing $r_i + 1$ of the i -th *kind* before $r_j + 1$ of the j -th kind is independent of the presence of the balls of other kinds and identical with P_{ij} where $i \neq j$ and

$$(14) \quad P_{ij} \equiv \psi_{(r_i, s_i, r_j, s_j)}.$$

Thus we see that the probability that $r_i + 1$ balls of the i -th kind be so drawn before $r_j + 1$ of the j -th kind, where $i \neq j = 1, \dots, k$ is *exactly* P_i defined by the relation

$$(15) \quad P_i \equiv \prod_{j \neq i}^k P_{ij} \equiv 2 \prod_{j=1}^{j \neq i} P_{ij}.$$

Arbitrarily, as in the case $k = 2$, we may apportion *individuals* among the k rival *treatments* by assigning to each T_i the portion, f_i , or making the chance of this assignment equal f_i , respectively. We may thus arbitrarily take $f_i = P_i$,

which may be calculated or we may use the machine, as we have seen that a unique answer is given at each turn just as in the special case, considered previously. Unlike that case, however, we are unable to state that P_i is the probability that T_i is the best of the k rivals; but its composition in (15) indicates that it may well serve the proposed purpose.

TABLE.

p_1	p_2	Total Trials ($n + n'$)	Trials of T_1 (n)	Failures of T_1 (r)	Approx. ($n \cdot p_2$)*
1	0	20	2, 1, 1, 1	2, 1, 1, 1	0
1	1/2	40	5, 9, 7, 5	5, 9, 7, 5	2, 4, 3, 2
1/2	0	40	6, 2, 3, 5	2, 2, 2, 2	0
3/4	1/4	100	3, 4	3, 3	1, 1
1	3/4	100	14, 10	14, 10	10, 7
3/4	1/2	100	23, 14	17, 11	8, 5
1/2	1/4	100	10, 13	5, 6	2, 3
1/4	0	100	4, 6	1, 1	0

YALE UNIVERSITY.

* *Expectation* of loss in the same n had T_2 been used.

ON A GENERALIZED TANGENT VECTOR.*

By H. V. CRAIG.

1. *Introduction.* The purpose of this paper is: (a) to prove that the left members, E_r , of the Euler equations associated with the function $F(x^1, \dots, x^n; dx^1/dt, \dots, dx^n/dt; \dots; d^m x^1/dt^m, \dots, d^m x^n/dt^m)$ and the quantities T_r , to be introduced, transform as the components of covariant tensors; (b) to make manifest certain points of similarity existing between T_r and the covariant tangent vector of Synge-Taylor geometry; and (c) to indicate a rôle that T_r might play in the development of a geometry based on F .

In Section 3 we develop certain formulas based on the rule for differentiating a product and in 4 apply them to establish by induction the covariance of E_r and T_r . These tensors are associated with the function F and the induction consists of proving that if for a given $F(x, \dots, d^m x/dt^m)$ the T_r and E_r related to $F(x, \dots, d^{m-1} x/dt^{m-1}, K)$ (K is a set of n constants) are tensors then the same may be asserted of the T_r and E_r corresponding to $F(x, \dots, d^{m-1} x/dt^{m-1}, d^m x/dt^m)$.

2. *Notation.* The symbolism to be employed in this paper is exhibited in the following table:

$$\begin{aligned} x'^r &= dx^r/dt; \quad x^{(p)r} = d^p x/dt^p; \quad {}_0C_0 = 1; \quad {}_mC_u \text{ is a binomial coefficient;} \\ \partial x^r/\partial y^i &= X_i^r = X_{(o)i}^{(o)r}; \quad \partial x^{(v)r}/\partial y^{(u)i} = X_{(u)i}^{(v)r}; \quad \partial F/\partial x^{(u)r} = F_{(u)r}; \\ \partial \bar{F}/\partial y^{(u)i} &= \bar{F}_{(u)i}; \quad T_r = \sum_{u=1}^m u(-1)^{u-1} F_{(u)r}^{..(u-1)}; \quad E_r = \sum_{u=0}^m (-1)^{u+1} F_{(u)r}^{..(u)}; \\ S_r &= F_{(o)r} + \sum_{u=2}^m (u-1)(-1)^{u-1} F_{(u)r}^{..(u)} - T_i\{^i_r\}. \end{aligned}$$

3. *Preliminary formulae.* The point transformation of $x^r = x^r(y)$ gives rise to the following equalities:

$$\begin{aligned} x'^r &= X_i^r y'^i; \quad x^{(m+1)r} = (X_i^r y'^i)^{(m)} = \sum_{u=0}^m {}_mC_u X^{(m-u)r} y'^{(u+1)i}; \\ F(x, x', \dots, x^{(m)}) &= \bar{F}(y, y', \dots, y^{(m)}); \quad X_{(m+1)j}^{(m+1)r} = X_j^r; \\ X_{(m)j}^{(m+1)r} &= {}_mC_{m-1} X_j^{(1)r} + {}_mC_0 X_j^{(1)r} = {}_{m+1}C_1 X_j^{(1)r}. \end{aligned}$$

This last relationship suggests the formula:

$$(1) \quad X_{(m-1)j}^{(m)r} = {}_mC_1 X_j^{(1)r}$$

* Presented to the Society, June 20, 1934.

which we shall now proceed to establish by induction, thus

$$\begin{aligned} X_{(m-l)j}^{(m+1)r} &= mC_{m-l-1}X_j^{(l+1)r} + \sum_{u=0}^l mC_u X_{(m-l)j}^{(m-u)r} y^{(u+1)i} \\ &= mC_{m-l-1}X_j^{(l+1)r} + \sum_{u=0}^l mC_u m-uC_{l-u}X_{ji}^{(l-u)r} y^{(u+1)i} \\ &= mC_{m-l-1}X_j^{(l+1)r} + mC_l \sum_{u=0}^l C_{l-u}X_{ji}^{(l-u)r} y^{(u+1)i} \\ &= m+1C_{l+1}X_j^{(l+1)r}. \end{aligned}$$

A second equality to be used in the sequel is as follows:

$$(2) \quad \sum_{u=1}^m u(-1)^{u-1} (F X_{(u)i}^{(m)r})^{(u-1)} = m(-1)^{m-1} F^{(m-1)} X_i^r.$$

To verify this we note that by virtue of (1) the left member of the foregoing may be written

$$\sum_{u=1}^m u(-1)^{u-1} mC_u \sum_{s=0}^{u-1} C_s F^{(s)} X_i^{(m-1-s)r}.$$

Now if s is not $m-1$ we have

$$\begin{aligned} \sum_{u=s+1}^m u(-1)^{u-1} mC_{m-u} C_{u-1} C_{u-1-s} &= m \sum_{u=s+1}^m C_{m-1-s} (-1)^{u-1} C_{u-s-1} \\ &= \pm m \sum_{l=0}^{m-s-1} (-1)^l C_l = 0, \end{aligned}$$

from which (2) follows.

4. *The vector character of T_r and E_r .* The covariance of T_r for m equal to one or two may be established* readily and so we pass on to the induction. Thus, if for any function $F(x, x', \dots, x^{(m-1)})$ T_r is a covariant vector, then we may write

$$\begin{aligned} \sum_{u=1}^m u(-1)^{u-1} \bar{F}_{(u)i}^{(u-1)} &= \sum_{u=1}^{m-1} u(-1)^{u-1} \bar{F}_{(u)r}^{(u-1)} X_i^r + \\ &+ \sum_{u=1}^m u(-1)^{u-1} (F_{(m)r} X_{(u)i}^{(m)r})^{(u-1)} \end{aligned}$$

from which we attain our conclusion by way of (2).

* See H. V. Craig, "On parallel Displacement in a non-Finsler space," *Transactions of the American Mathematical Society*, vol. 33 (1931), p. 133.

† The assumption that the T_r associated with $F(x, x', \dots, x^{(m-1)}, k)$ is a tensor implies that for $F = F(x, x', \dots, x^{(m-1)}, k)$

$$\sum_{u=1}^{m-1} u(-1)^{u-1} \sum_{p=u}^{m-1} (F_{(p)r} X_{(u)i}^{(p)r})^{(u-1)} = \sum_{u=1}^{m-1} u(-1)^{u-1} \bar{F}_{(u)r}^{(u-1)} X_i^r$$

and from the nature of this reduction it follows that the same simplification can be made if k is replaced with $x^{(m)}$.

This accomplished there remains to be proved that the left members of the Euler equations transform according to a tensor law, or more explicitly that $\bar{E}_i = E_r X_i^r$. Again we employ mathematical induction, thus

$$\sum_{u=0}^m (-1)^{u+1} \bar{F}_{(u)i}^{..(u)} = \sum_{u=0}^{m-1} (-1)^{u+1} F_{(u)r}^{..(u)} X_i^r + \sum_{u=0}^m (-1)^{u+1} (F_{(m)r} X_{(u)i}^{(m)r})^{(u)}.$$

By expanding the last term of the foregoing and evaluating the derivatives of $x^{(m)r}$ by means of (1) we obtain the expression

$$\sum_{u=0}^m (-1)^{u+1} \sum_{s=0}^u {}_u C_s F_{(m)r}^{..(s)} {}_m C_u X_i^{(m-s)r}$$

which reduces to $(-1)^{m+1} F_{(m)r}^{..(m)} X_i^r$ since for s not m

$$\sum_{u=s}^m (-1)^{u+1} {}_m C_u {}_u C_s$$

is zero.

5. *Certain generalized geometries.* A metric manifold such that the arc length of a curve C ($C: x^r = x^r(t)$) is given by the integral $\int F(x^1, \dots, x^n; x'^1, \dots, x'^n) dt$ is called a Finsler space. The function F is among other things assumed to satisfy the identity $x'^r F_{(1)r} = F$; this insures the invariance under parameter change of the integral $\int F dt$. J. L. Synge and, independently, J. H. Taylor have investigated the geometry of a Finsler space having for its metric tensor the quantities f_{rs} ($2f_{rs} = F^2_{(1)r(1)s}$). As an immediate consequence of the identity $x'^r F_{(1)r} = F$ they derive that $x'^s f_{rs} = F F_{(1)r}$. Consequently if the parameter is the Finsler arc (in this case F maintains the value unity along the curve in question) the quantities x'^r , $F_{(1)r}$ are said to be the contravariant and covariant descriptions of the unit tangent vector. One of the salient properties of this geometry is that the auto-parallel curves $\theta x^r = 0$ * coincide with the extremals associated with F . Likewise, it may be proved easily that $\theta F_{(1)r} = E_r$ ($E_r = F'_{(1)r} - F_{(0)r}$). Furthermore, $x'^r \theta F_{(1)r} = 0$ and so the vector θF_{1r} may be regarded as the covariant principal normal vector.

Spaces involving metric tensors whose components are functions of not only x and x' but of higher derivatives as well were first investigated by Akitsugu Kawaguchi. Accordingly, we shall refer to the manifold associated with $\int F(x, x', \dots, x^m) dt$ as a Kawaguchi space. Incidentally, a Euclidean

* For a discussion of Synge-Taylor geometry including the θ process reference may be made to J. H. Taylor, "A generalization of Levi-Civita's parallelism and the Frenet formulas," *Transactions of the American Mathematical Society*, vol. 27 (1925), p. 255 or J. L. Synge, "A generalization of the Riemannian line-element," *ibid.*, p. 61.

plane may be made the bearer of a Kawaguchi space in the following manner. Let there be given the set of all plane curves of class $C^{(m)}$, $x = x(t)$, $y = y(t)$ together with the set of normals to the x, y plane and let each of these curves be warped into the corresponding space curve; $x = x(t)$, $y = y(t)$, $z = \int_0^t (F^2(x, y; x', y'; \dots; x^{(m)}, y^{(m)}) - (x'^2 + y'^2))^{1/2} dt$. Obviously, the length of arc of the part of one of these curves that joins the normals at $P_1(x_1, y_1)$, $P_2(x_2, y_2)$ is given by the integral $\int_{P_1}^{P_2} F dt$ taken along the base curve.

In addition to the evident requirements as to differentiability etc., we shall assume in what follows that F satisfies two conditions, namely: (a) F is positive along each regular curve; (b) $F dt$ is invariant in functional form under an admissible parameter transformation.

6. *The vector T_r .* Obviously, if m is one, T_r is $F_{(1)r}$ and so our "tangent" vector is a generalization of the covariant tangent vector of Synge-Taylor geometry. Furthermore, in this case it is well known that (b) implies the identity $x'^r T_r = F$ and, as a matter of fact, this same implication has been established for $m = 2$.* Thus we are led to consider the situation in general.

As a preliminary we shall demonstrate that

$$(3) \quad \sum_{v=1}^{m+1} (-1)^{v-1} {}_m C_{v-1} [x^{(m+2-v)} F]^{(v-1)} = (-1)^m x' F^{(m)}$$

is an identity.

Proof. By the rule for differentiating products, we have

$$[x^{(m+2-v)} F]^{(v-1)} = \sum_{w=0}^{v-1} {}_{v-1} C_w x^{(m+1-w)} F^{(w)}.$$

Consequently, the coefficient of $x^{(m+1-w)} F^{(w)}$ in (3) is $\sum_{v=w+1}^{m+1} (-1)^{v-1} {}_m C_{v-1} \cdot {}_{v-1} C_w$, which, by virtue of the equality ${}_m C_{v-1} {}_{v-1} C_w = {}_m C_{m-w} {}_{m-w} C_{v-1-w}$ and the substitution $v = u + w + 1$ may be written ${}_m C_{m-w} \sum_{u=0}^{m-w} (-1)^{u+w} {}_{m-w} C_u$. But, by a well known property of the binomial coefficients this last expression is $(-1)^m \delta_w^m$ and the lemma is established. This accomplished, we turn to the

THEOREM. *A necessary condition for the invariance of functional form of $F(x, x', \dots, x^{(m)}) dt$ under a parameter transformation is $x'^r T_r = F$.*

* See H. V. Craig, "On parallel displacement in a non-Finsler space," *loc. cit.*, p. 133.

Proof. If F has the invariant property in question and T is any function of t then

$$(FT)' = \sum_{u=0}^m (x'^r T)^{(u)} F_{(u)r}^*.$$

From this by setting $T = t, t^2/2!$ etc. successively, we derive that

$$\sum_{u=v}^m u C_v x^{(u-v+1)r} F_{(u)r} = \delta_v^1 F \quad (v = 1, 2, \dots, m).$$

Designating the left member of this equation with L_v we find by direct calculation, for small values of m) that $\sum_{v=1}^m (-1)^{v-1} v L_v^{(v-1)}$, which is obviously F , reduces identically in $F^{(v)}$ to $x'^r T_r$. If we assume that this reduction takes place for a given value of m and, in the case $m+1$, represent $T_r = (-1)^m (m+1) F_{m+1,r}^{(m)}$ with T_r' , then we may write

$$\begin{aligned} \sum_{v=1}^{m+1} (-1)^{v-1} v L_v^{(v-1)} &= \sum_{v=1}^m (-1)^{v-1} v \sum_{u=v}^m u C_v [x^{(u-v+1)r} F_{(u)r}]^{(v-1)} \\ &\quad + \sum_{v=1}^{m+1} (-1)^{v-1} v_{m+1} C_v (x^{(m+2-v)r} F_{(m+1)r}^{(v-1)}). \end{aligned}$$

But the first term in the right member of the foregoing is by assumption $x'^r T_r'$, while the second can be put in the form

$$(m+1) \sum_{v=1}^{m+1} (-1)^{v-1} {}_m C_{m+1-v} (x^{(m+2-v)r} F_{(m+1)r}^{(v-1)})$$

which by (3) reduces identically to $(-1)^m (m+1) x'^r F_{m+1,r}^{(m)}$, and hence the theorem follows.

As a consequence of this we can so select the parameter that $x'^r T_r$ will maintain the value unity along any prescribed regular curve. Also, if we were to choose the quantities $F_{(m)r(m)s} + T_r T_s$ as the components of the metric tensor f_{rs} then, because $x'^r F_{(m)r(m)s} = 0$, we would have $x'^s f_{rs} = T_r$, and, with a properly selected parameter, $x'^r x'^s f_{rs} = 1$; $T_r T_s f^{rs} = 1$.

With regard to possible future developments based on T_r , we note that an obvious consequence of the definitions of T_r and S_r is the following: if $\{j\}$ is any two index connection \dagger then the extremal curves associated with

* See Adolph Kneser, *Lehrbuch der Variationsrechnung*, Braunschweig (1900), p. 195.

\dagger I. e. an object which transforms as $x'^s \{^j_{sr}\}$, see L. P. Eisenhart, *Riemannian Geometry* (1926), p. 19. For a most general connection reference may be made to A. Kawaguchi, "Die Differentialgeometrie in der verallgemeinerten Mannigfaltigkeit,"

F are those for which the vectors θT_r ($\theta T_r = T'_r - T_j \{j\}$) and S_r coincide. Should F and the connection be such that S_r is zero then the extremal curves may be characterized as auto-parallel and in this case we may conclude from (b) that $x^r \theta T_r$ vanishes.† As a matter of fact such connections may be constructed. For if $\{j\}$ is any two index connection and the generalized arc the parameter, then the quantities $\{j\}^*$ defined by $\{j\}^* = \{j\} + x'^j S_r$ also constitute a connection. Evidently, this connection is such that the associated S^*_r vanishes, for S^*_r may be written $S_r - T_j x'^j S_r$.

TEXAS UNIVERSITY.

Rendiconti di Palermo, tomo 56 (1932), pp. 245-276; also see H. Hombu, "On a non-Finsler metric space," *Tohoku Mathematical Journal*, vol. 37 (1933), pp. 190-198.

† See H. V. Craig, "On the solution of the Euler equations for their highest derivatives," *Bulletin of the American Mathematical Society*, vol. 36 (1930), p. 560.



CYCLOTOMY, HIGHER CONGRUENCES, AND WARING'S PROBLEM II.¹

By L. E. DICKSON.

PART 2. THE WARING PROBLEM FOR POLYNOMIAL SUMMANDS.

29. *Introduction and summary.* In 1770, E. Waring conjectured that every positive integer N is a sum of 9 integral cubes ≥ 0 , also that N is a sum of 19 fourth powers, etc. Hardy and Littlewood proved that every sufficiently large N is a sum of

$$(160) \quad s = \left(\frac{1}{2}k - 1\right)2^{k-1} + k + 5 + \xi_k$$

integral k -th powers ≥ 0 , where ξ_k is the greatest integer \leq the quotient of $(k-2) \log 2 - \log k + \log(k-2)$ by $\log k - \log(k-1)$. Except for very small values of k , ξ_k is quite small compared to (160); for example, $\xi_{10} = 50$, $\xi_{28} = 493$.

Waring conjectured also that every N is a sum of a limited number of values of a polynomial in x of degree k . In precise form, this was proved by E. Kampke.² But neither writer gave any information as to the number of values needed. For $k=3$, 9 values suffice.³ For $k \geq 4$ the analytic part of the proof that s values of a polynomial suffice for a large N has been made by Miss Humphreys.⁴ We here treat the second part of the proof, viz., that if A is any integer and p is any prime not dividing k , then A is congruent modulo p to a sum of n values of the polynomial, where $n < s$ in (160). We find that

k	3	4	5	6	7	8	9	10
n	4	6	12	24	48	72	144	216
s	9	19	41	87	192	425	949	2113.

If k is one of the even numbers 6, 8, \dots , 18, then

$$n = n(k) \leq 8 \cdot 3^{\frac{1}{2}k-2},$$

which is less than the first term of (160) since $3 < 4$.

¹ Part I of this paper appeared in the current volume of this JOURNAL, pp. 391-424.

² *Mathematische Annalen*, Bd. 83 (1921), pp. 85-112.

³ Dickson, *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 1-12, 739-741; R. D. James, *American Journal of Mathematics*, vol. 56 (1934), pp. 303-315.

⁴ *Duke Mathematical Journal*, vol. 1 (1935).

⁵ For the case when p divides k , see § 45.

For any odd k , $n(k) \leq 2n(k-1)$, whence $n(k)$ is much less than the first term of (160) when $n=7, 9, \dots, 19$. If we write $s = S + \xi_k$, then

k	20	22	24	26
n	384,912	1,154,736	57,736,800	173,210,400
S	4,718,617	20,971,547	92,274,717	402,653,215.

Hence $n < s$ for $k < 28$. For $n \geq 28$, s exceeds a billion, and the actual value of s is of slight interest beyond the fact that there exists an s (Kampke).

30. *Normal polynomials.* We may exclude polynomials $g(x)$ whose values for integers x are all multiples of p , since integers not multiples of p are not represented as a sum of values of $g(x)$. The true Waring problem relates to summands g/p and not to summands g .

By the degree of $g(x)$ modulo p we mean the exponent of the highest power of x whose coefficient c is prime to p . We seek n such that

$$(161) \quad A \equiv \sum_{i=1}^n g(x_i) \pmod{p}$$

has integral solutions x_i for every integer A . We desire that the same n shall serve not only for every A , but for every polynomial $g(x)$ of given degree k modulo p . We write $n = n(k) = n(k, p)$.

Determine d by $cd \equiv 1 \pmod{p}$. Then dA ranges with A over a complete set of residues modulo p . Hence the problem for (161) reduces to that for $dg(x)$, whose c is $\equiv 1 \pmod{p}$. If k is prime to p , we take $x = X + z$ and choose z so that the coefficient of X^{k-1} is divisible by p . If C is the constant term of $g(x)$, write $g = H + C$. Then (161) is equivalent to $A - nC \equiv \Sigma H$.

THEOREM 13. *When k is not divisible by p , the problem for (161) reduces to the like problem for a NORMAL polynomial whose leading coefficient is unity, while the coefficient of x^{k-1} and the constant term are both zero.*

LEMMA 1. *If neither r nor s is divisible by $p > 2$ and if A is any integer, there exist solutions of*

$$rx^2 + sy^2 \equiv A \pmod{p}.$$

This special case of Theorem 6 is evident since each of rx^2 and $A - sy^2$ takes $1 + \frac{1}{2}(p-1)$ incongruent values and hence have a common value.

When $k=2$ the only normal polynomial is x^2 . Then Lemma 1 with $r = s = 1$ gives

$$(162) \quad n(2, p) = 2, \quad p > 2.$$

31. *Odd k .* Let k be not divisible by p . By choice of z , $g(x+z)$ becomes

$$f(x) = \sum_{i=0}^k c_i x^{k-i}, \quad c_1 \text{ not divisible by } p.$$

$$f(x) + f(-x) = 2c_1 x^{k-1} + 2c_3 x^{k-3} + \dots = H(x).$$

Take $p > 2$. Then $2c_1$ is prime to p . By definition, any A is congruent to a sum of $n(k-1, p)$ values of $H(x)$.

THEOREM 14. *If k is odd and not divisible by $p > 2$, then*

$$n(k, p) \leq 2n(k-1, p).$$

COROLLARY. $n(3, p) \leq 4$ if $p > 3$.

32. *Case $k=4$.* We employ the fact that the sum of the fourth powers of $a+d$, $a-d$ and $-2a$ is $2t^2$, the sum of their squares is $2t$, and their sum is zero, where $t = 3a^2 + d^2$. For $p > 2$ every normal polynomial of degree 4 is of the form $f(x) = x^4 + 2ux^2 + vx$. Hence

$$2t^2 + 2t \cdot 2u = f(a+d) + f(a-d) + f(-2a).$$

The left member is $2(y^2 - u^2)$, where $y = t + u$. Employ also a second such identity and add the two. Hence $N = 2(y^2 + z^2 - 2u^2)$ is a sum of six values of $f(x)$.

Take $p > 3$ and apply Lemma 1. Thus integers y and z may be chosen so that N takes any assigned value modulo p . For the t determined by y , $3a^2 + d^2 \equiv t \pmod{p}$ is solvable. Similarly for τ determined by z , $3\alpha^2 + \delta^2 \equiv \tau \pmod{p}$ is solvable. This proves

$$(163) \quad n(4, p) \leq 6 \quad \text{if } p > 3.$$

Hence by Theorem 14,

$$(164) \quad n(5, p) \leq 12 \quad \text{if } p > 5.$$

33. LEMMA 2. *For every integer A , there exist ⁶ integral solutions of*

$$(165) \quad \sum_{i=1}^r h_i^k \equiv A \pmod{p},$$

where r is the g. c. d. of k and $p-1$.

Taking $A = -1$, $k_{r+1} = 1$, we obtain

⁶ Landau, *Vorlesungen über Zahlentheorie*, I (1927), p. 290. The proof is quite elementary.

LEMMA 3. For $K = r + 1$, there exist solutions of

$$(166) \quad \sum_{i=1}^K h_i x_i \equiv 0, \text{ not every } h_i \equiv 0 \pmod{p}.$$

34. *Even k .* Let k be not divisible by $p > 2$. As in § 30, it suffices to consider a polynomial of the form $f(x) = x^k + x^{k-1} + \dots$. By (166),

$$(167) \quad P(y) = \sum_{i=1}^K f(h_i y) \equiv Cy^{k-1} + \dots \pmod{p}, \quad C = \sum_{i=1}^K h_i^{k-1}.$$

If C is not divisible by p , we have the result desired. Next, let $C \equiv 0$. By (166) a certain h_j is prime to p . Let $Q(y)$ be derived from $P(y)$ by changing the sign of h_j . If also the leading coefficient of $Q(y)$ is divisible by p , evidently $2h_j^{k-1} \equiv 0$, $k_j \equiv 0 \pmod{p}$, contrary to hypothesis.

THEOREM 15. Let k be even and not divisible by the odd prime p . Choose $K (\leq r + 1)$ so that (166) is solvable. Then

$$n(k, p) \leq Kn(k - 1, p).$$

35. *Lemmas, chiefly on congruences.*

LEMMA 4. If q is prime to $p - 1$, every integer is congruent to a q -th power modulo p .

Since there are integral solutions of $v(p - 1) + 1 = uq$,

$$x \equiv x(x^{p-1})^v \equiv (x^u)^q \pmod{p}.$$

LEMMA 5. Let r be the g. c. d. of k and $p - 1$. If each a_i is prime to p ,

$$(168) \quad a_1 x_1^k + \dots + a_s x_s^k \equiv c \pmod{p}$$

has the same number of solutions as

$$(169) \quad a_1 y_1^r + \dots + a_s y_s^r \equiv c \pmod{p}.$$

Consider any solution of $a_1 z_1 + \dots + a_s z_s \equiv c \pmod{p}$. For ($i = 1, \dots, s$), we shall prove that $x_i^k \equiv z_i$ and $y_i^r \equiv z_i \pmod{p}$ have the same number of roots. This is evident unless z_i is prime to p . Then

$$k \text{ Ind } x_i \equiv \text{Ind } z_i \pmod{p - 1}$$

has no root or r roots x_i according as $\text{Ind } z_i$ is not or is divisible by r . The same is true for $r \text{ Ind } y_i \equiv \text{Ind } z_i$.

From Theorem 6 and Lemma 5 we obtain

LEMMA 6. If $p \equiv -1 \pmod{4}$ and if q is prime to $p - 1$, there are exactly $p + 1$ solutions of

$$x^k + y^k \equiv -1 \pmod{p}, \quad k = 2^m q, \quad m \geq 1.$$

The conditions are satisfied if $p \equiv -1 \pmod{12}$, $q = 3^n$.

LEMMA 7. If $p \equiv 1 \pmod{4}$ and if q is prime to $\frac{1}{2}(p-1)$, there are exactly $p-1$ solutions of

$$x^{2q} + y^{2q} \equiv -1 \pmod{p}.$$

LEMMA 8. If $p > 2$, there are at most km simultaneous solutions of

$$(170) \quad h^k + H^k \equiv -1, \quad h^m + H^m \equiv -1 \pmod{p}, \quad m \neq k.$$

Let d be the g. c. d. of k and m . Comparing $(-H^k)^{m/d}$ and $(-H^m)^{k/d}$, we get

$$(171) \quad (h^k + 1)^{m/d} \pm (h^m + 1)^{k/d} \equiv 0,$$

which is not identically $\equiv 0$. It has at most km/d roots. Since d is a linear combination of k , m , we see from (170) that H^d is congruent to a polynomial in h .

LEMMA 9. $n(k, p^e) \leq p^e - 1$.

We exclude polynomials $f(x)$ all of whose values are multiples of p . As in § 30, we may assume that the constant term of $f(x)$ is zero. Let v denote a value prime to p of $f(x)$.

I. If A is any integer prime to p , $tv \equiv A \pmod{p^e}$ has a solution t , $1 \leq t < p^e$. Thus A is congruent to a sum of t (equal) values of $f(x)$.

II. If $A \equiv 0 \pmod{p^e}$, then $A \equiv f(0) \pmod{p^e}$.

III. Let $A = p^m a$, $1 \leq m < e$, a prime to p . By I,

$$a = S + zp^{e-m}, \quad S = \text{sum of } p^{e-m} - 1 \text{ values of } f(x).$$

Multiply by p^m . Hence A is congruent modulo p^e to

$$p^m S = \text{sum of } p^m(p^{e-m} - 1) < p^e - 1 \text{ values of } f(x).$$

36. Case $k = 6$. We shall prove that $n(6, p) \leq 24$ if $p > 3$.

I. $p \equiv 1 \pmod{4}$. Then $h^2 \equiv -1 \pmod{p}$ is solvable. Hence $h^6 + 1 \equiv 0$. Thus $K = 2$ in (166) and $n(6, p) \leq 2 \cdot 12$ by Theorem 15 and (164) if $p > 5$. For $p = 5$, use Lemma 9.

II. $p \equiv -1 \pmod{12}$. By Lemma 6, there are exactly $p+1$ solutions of

⁷ If m/d and k/d are both odd, there are at most $2\delta + km - 3\delta m$ simultaneous solutions, where δ is the number of roots of $z^d \equiv -1 \pmod{p}$.

$$(172) \quad h_1^6 + h_2^6 + h_3^6 \equiv 0 \pmod{p}, \quad h_3 = 1.$$

If $p+1 > 24$, Lemma 8 shows that (172) has a solution for which $h_1^4 + h_2^4 \not\equiv -1$, and one for which $h_1^2 + h_2^2 \not\equiv -1 \pmod{p}$. But if $p = 11$ or 23 , $n(6, p) \leq 22$ by Lemma 9.

Consider a normal polynomial

$$(173) \quad f(x) = x^6 + c_4x^4 + \cdots + c_1x.$$

Then

$$(174) \quad P(y) = \sum_{i=1}^3 f(h_i y) \equiv \sum_{j=1}^4 C_j M_j y^j \pmod{p}, \quad M_j = \sum_{i=1}^3 h_i^j.$$

If $P(y)$ is not identically $\equiv 0 \pmod{p}$, $n(6, p) \leq 3n(4, p) \leq 18$. Next, let $P(y)$ be identically $\equiv 0$ for all solutions of (172). Since (172) holds also when $h_3 = -1$, there are solutions with $M_3 \not\equiv 0$, whence $c_3 \equiv 0$. Similarly, $M_1 \not\equiv 0$, $c_1 \equiv 0$. We saw that there are solutions with $M_2 \not\equiv 0$, $M_4 \not\equiv 0$, whence $c_2 \equiv 0$, $c_4 \equiv 0$. Hence $f(x) \equiv x^6$ and $n(6, p) = 2$ by Lemma 2 or Lemmas 1 and 4.

LEMMA 10. *Except for $p = 7, 31, 67, 79, 139, 223$, there exist solutions of $h^6 + H^6 \equiv -1 \pmod{p}$, if $p \equiv 7 \pmod{12}$.*

Since $p = 6f + 1$, f is odd and the number N of solutions is $36(0, 3)$ by Theorem 5. By § 19,

$$N = p + 1 + 16A, \quad N = p + 1 + 10A \pm 12B, \quad p = A^2 + 3B^2,$$

according as 2 is or is not a cubic residue of p . The sign of A was there chosen so that $A \equiv 4 \pmod{6}$. By Theorem 7, B is a multiple $3y$ of 3 in the first case; but in the second case, B is prime to 3 and we may choose the sign so that $\pm B \equiv A \pmod{3}$.

Let $N = 0$. Eliminate p . In the first case,

$$7 = \left\{ \frac{1}{3}(A + 8) \right\}^2 + 3y^2 = 4 + 3, \quad A = -2 \text{ or } -14, \quad B^2 = 9, \quad p = 31 \text{ or } 223.$$

In the second case, p and 37 are the products of $A \pm B\sqrt{-3}$ and $5 - 2\sqrt{-3}$ by their conjugates, whence, by multiplication,

$$37p = X^2 + 3Y^2, \quad X = 5A \pm 6B \equiv 2 \pmod{6}, \quad Y = -2A \pm 5B,$$

and $Y = 3w$, $X + 37 = 3v$, v odd. If $N = 0$, $p = -1 - 2X$,

$$148 = v^2 + 3w^2 = 121 + 3 \cdot 9 \text{ or } 1 + 3 \cdot 49,$$

whence $p = 7, 139$ or $67, 79$.

III. $p \equiv 7 \pmod{12}$. First, exclude the six p 's in Lemma 10. Then (172) is solvable. There exists an integer e belonging to the exponent 6 modulo p . Hence (172) holds also when h_1 is replaced by eh_1 . Hence (172) has solutions for which $M_j \not\equiv 0$, $c_j \equiv 0$ ($j=1, \dots, 4$ in turn) in (174).

We have the following solutions of (166) with $K=4$, $k=6$.

$$\begin{aligned} p=67, & \quad 1+1+1+2^6 \equiv 0; \\ p=79, & \quad 1+1+10+67 \equiv 0, \quad 10 \equiv g^6, \quad 67 \equiv g^{54}, \quad g=29; \\ p=139, & \quad 1+1+6+131 \equiv 0, \quad 6 \equiv g^{30}, \quad 131 \equiv g^{12}, \quad g=92; \\ p=223, & \quad 1+4+8+210 \equiv 0, \quad 2 \equiv 10^{18}, \quad 210 \equiv 10^{48}. \end{aligned}$$

For each such p , $n(6, p) \leq 4 \cdot 6$.

LEMMA 11. If $p \equiv 1 \pmod{k}$ and if every integer is congruent modulo p to a sum of s k -th powers, then $n(k, p) \leq sk$.

There exist k roots h_i of $h^k \equiv 1$, whence $\sum h_i^j \equiv 0 \pmod{p}$ for $1 \leq j < k$ by Newton's identities. We may take $f(x) = x^k + \dots$. Then

$$\sum_{i=1}^k f(h_i y) \equiv ky^k \pmod{p}.$$

Since r is now k in Lemma 2, we have

LEMMA 12. If $p \equiv 1 \pmod{k}$, $n(k, p) \leq k^2$.

For $k=6$, $p=31$, we find that $s=4$ in Lemma 11, whence $n(6, 31) \leq 24$.

For $p=7$, apply Lemma 9.

THEOREM 16. If $p > 3$, $n(6, p) \leq 24$.

37. Case $k=8$. Proof that $n(8, p) \leq 72$.

I. $p \equiv -1 \pmod{4}$. By Lemma 6, there are exactly $p+1$ solutions of

$$(175) \quad h^8 + H^8 \equiv -1 \pmod{p}.$$

By Lemma 8, (175) has at most 48 solutions in common with one of $h^6 + H^6 \equiv -1$, $h^4 + H^4 \equiv -1$, $h^2 + H^2 \equiv -1$. Hence if $p+1 > 48$ there is a solution of (175) for which $M_6 \not\equiv 0$, one for which $M_4 \not\equiv 0$, one for which $M_2 \not\equiv 0$, where the notations refer to (174) with ($j=1, \dots, 6$), whence $n(8, p) \leq 3n(6, p) \leq 72$. For $p+1 \leq 48$, we have $n(8, p) \leq 46$ by Lemma 9.

II. $p \equiv 1 \pmod{8}$. Then $n(8, p) \leq 64$ by Lemma 12.

LEMMA 13. Let $p = 4f + 1 = x^2 + 4y^2$, $x \equiv 1 \pmod{4}$. The number N of solutions of $h^4 + H^4 \equiv -1 \pmod{4}$ is $p - 3 - 6x$ if f is even, but $p + 1 - 6x$ if f is odd.

For, by Theorems 2 and 5, $N = 8 + 16(00)$ or $16(02)$. Apply (52) and (56).

III. $p \equiv 5 \pmod{8}$. If in Lemma 13, $p + 1 - 6x = 0$, eliminate p from $x^2 + 4y^2 = p$. Thus $z^2 + y^2 = 2$, $z = \frac{1}{2}(x - 3)$. Hence $z = \pm 1$, $x = 1$ or 5 , $p = 5$ or 29 .

Let $p \neq 5$, $p \neq 29$. Then $h^4 + H^4 \equiv -1$ has solutions. By Lemma 5 it has the same number of solutions as (175). Thus

$$(176) \quad h_1^8 + h_2^8 + h_3^8 \equiv 0 \pmod{p}$$

has solutions with h_3 prime to p . There exists an integer e belonging to the exponent 4 modulo p . We have (174) with $(j = 1, \dots, 6)$. Let the new (174) be identically $\equiv 0 \pmod{p}$ for all solutions of (176). As below (174), $f(x)$ involves only even powers of x . We do not alter (176) if we replace h_3 by eh_3 . If the old M_2 is $\equiv 0$, the new M_2 is $\not\equiv 0$, whence $c_2 \equiv 0$. Similarly $c_6 \equiv 0$. Hence $f(x) \equiv x^8 + c_4x^4$. Employ

$$4(a + b)^4 + 4(a - b)^4 + (2a)^4 + (2b)^4 = 24(a^2 + b^2)^2.$$

The corresponding sum of eighth powers is $8S$, where

$$S = 33(a^8 + b^8) + 28(a^6b^2 + a^2b^6) + 70a^4b^4.$$

We may take $a = 1$, $b^2 \equiv -1 \pmod{p}$. Then $S \equiv 80$. Hence there exist solutions of

$$\sum_{i=1}^{10} h_i^4 \equiv 0, \quad \sum h_i^8 \equiv 8 \times 80 \not\equiv 0 \pmod{p}.$$

Then $f(h_i y) \equiv 640y^8$. In Lemma 2, r is now 4. Hence every integer is congruent to a sum of 4×10 values of $x^8 + c_4x^4$.

For $p = 29$, $n(8, p) \leq 28$ by Lemma 9.

THEOREM 17. $n(8, p) \leq 72$.

38. Case $k = 2q$, q a prime > 3 .

I. $p \equiv 1 \pmod{q}$. Then $n(k, p) \leq k^2$ by Lemma 12.

II. $p \not\equiv 1 \pmod{q}$. By Lemmas 6, 7, there are exactly $p \pm 1$ solutions of

$$(177) \quad h^k + H^k \equiv -1 \pmod{p},$$

according as $p \equiv \mp 1 \pmod{4}$. Apply Lemma 8 with $m < k$. Thus $n(k, p) \leq 3n(k-2, p)$ if $p \pm 1 > k(k-2)$.

THEOREM 18. If $p \geq 7$, $n(10, p) \leq 216$.

This was proved if $p \pm 1 > 80$. But if ≤ 79 , apply Lemma 9.

39. Case $k = 12$. To prove that $n(12, p) \leq 648$.

I. $p \equiv 1 \pmod{12}$. By Lemma 12, $n(12, p) \leq 144$.

II. $p \equiv -1 \pmod{12}$. By Lemma 6,

$$(178) \quad h^{12} + H^{12} \equiv -1 \pmod{p}$$

has exactly $p+1$ solutions. If $p+1 > 120$, Lemma 8 shows that $n(12, p) \leq 3n(10, p) \leq 648$. If $p+1 \leq 120$, apply Lemma 9.

III. $p \equiv 5 \pmod{12}$. Since every integer is congruent to a cube (Lemma 4), the number N of solutions of (178) is the same as the number of solutions of

$$(179) \quad z^4 + w^4 \equiv -1 \pmod{p},$$

which is true also by Lemma 5. Here $p \equiv 5$ or $17 \pmod{24}$.

IV. $p \equiv 5 \pmod{24}$. By Lemma 13, $N = p+1-6x$, $x \equiv 1 \pmod{4}$, $x^2 + 4y^2 = p$. Thus $N \equiv 0 \pmod{24}$. Since -2 is a quadratic non-residue of every prime $p \equiv 5 \pmod{8}$, $z^4 \not\equiv w^4$ in (179). Also, $w^4 \equiv -1$ implies $w^8 \equiv 1$, $w^{p-1} \equiv 1$, $w^4 \equiv 1$, whence $z \not\equiv 0$, $w \not\equiv 0$ in (179). But $z^4 \equiv 1$ has four roots. Hence N is a multiple of $2 \times 4 \times 4$. Hence N is divisible by 96.

If $N > 12 \times 10$, Lemma 8 gives $n(12) \leq 3n(10)$. It remains to treat $N = 0$, $N = 96$. By III, § 37, $N = 0$ requires that $p = 5$ or 29 . If $N = 96$, eliminate $p = 6x + 95$ from $x^2 + 4y^2 = p$. Hence

$$26 = v^2 + y^2, \quad v = \frac{1}{2}(x-3) = \text{odd}, \quad v = \pm 5, \pm 1,$$

$p = 53, 101, 173(125)$. Apply Lemma 9.

V. $p \equiv 17 \pmod{24}$. By Lemma 13, $N = p-3-6x$. Hence $N \equiv 8 \pmod{24}$. Since there exists a number e belonging to the exponent 8, $e^4 \equiv -1 \pmod{4}$ has four roots. Thus (179) has eight solutions with $z \equiv 0$ or $w \equiv 0$, and hence has $N-8$ solutions z, w both prime to p . If the quadratic residue 2 is a residue of a fourth power, there are four solutions of $2z^4 \equiv -1$ and hence 16 solutions of (179) with $z^4 \equiv w^4$, whence $N-24$ is a multiple of $2 \times 4 \times 4$. This with $N \equiv 8 \pmod{24}$ gives $N \equiv 56 \pmod{96}$. Next let $z^4 \not\equiv w^4$, then $N-8$ is a multiple of 32. Hence $N \equiv 8 \pmod{96}$.

By Lemma 8, it remains to treat $N \leq 120$, whence $N = 8, 56$ or 104 . Eliminate p and write $v = \frac{1}{2}(x-3) = \text{odd}$.

If $N = 8$, $5 = v^2 + y^2$, $v = \pm 1$, $p = 17$ or 41 .

If $N = 56$, $17 = v^2 + y^2$, $v = \pm 1$, $p = 89$, or $53 \not\equiv 17 \pmod{24}$.

If $N = 104$, $29 = v^2 + y^2$, $v = \pm 5$, $p = 65$ or 185 (not primes).

For these primes apply Lemma 9.

VI. $p \equiv 7 \pmod{12}$. By Lemma 5, the number N of solutions of (178) is the same as that of $x^6 + y^6 \equiv -1$. If $N = 0$, Lemma 10 gives $p \leq 223$, whence Lemma 9 applies. Let $N > 0$. Since there exists an integer belonging to the exponent 6 modulo p , the usual proof gives $n(12) \leq 3n(10)$ unless $f(x) = x^{12} + cx^6$.

Suppose that (178) has a solution in common with

$$(180) \quad h^6 + H^6 \equiv -1 \pmod{p}.$$

Elimination of h^6 gives $H^{12} + H^6 + 1 \equiv 0$, whence $H^6 \not\equiv 1$. Thus the g. c. d. of the exponents in $H^{18} \equiv 1$, $H^{p-1} \equiv 1$ must exceed 6. Hence $p \equiv 19 \pmod{36}$. Conversely, $H^{12} + H^6 + 1 \equiv 0$ then has twelve roots. Write $h = tH^2$, $t^6 \equiv 1$. Then (180) holds and there are exactly 72 simultaneous solutions of (178) and (180).

It therefore remains only to treat the case $N = 72$. By the results below Lemma 10, $N = p + 1 + 16A$ or $p + 1 + 2X$. Eliminate p . In the first case, $15 = z^2 + 3y^2$, $z = \frac{1}{3}(-A-8)$, which is impossible. In the second case, $37 \cdot 12 = v^2 + 3w^2$. Thus $v = 3u$, $148 = 3u^2 + w^2$. The only solutions are $(u^2, w^2) = (9, 121)$, $(16, 100)$, $(49, 1)$. The p s are 19, 199, 271, 91 and 217 (factors 7), $73 \not\equiv 19 \pmod{36}$. Apply Lemma 8.

THEOREM 19. If $p \geq 7$, $n(12, p) \leq 648$.

40. Case $k = 14$. If $p \pm 1 > 14 \times 12 = 168$, § 38 applies. If $p \pm 1 \leq 168$, Lemma 9 applies. Hence

THEOREM 20. If $p \geq 7$, $n(14, p) \leq 1944$.

41. Case $k = 16$. To prove $n(16, p) \leq 5832$.

I. $p \equiv 1 \pmod{16}$. By Lemma 12, $n(16, p) \leq 256$.

II. $p \equiv -1 \pmod{4}$. By Lemma 6, with $q = 1$, there exist exactly $p + 1$ solutions of

$$(181) \quad h^{16} + H^{16} \equiv -1 \pmod{p}.$$

By Lemma 8, $n(16, p) \leq 3n(14, p)$ unless $p + 1 \leq 224$, and then Lemma 9 applies.

III. $p \equiv 5 \pmod{8}$. By Lemma 5, (181) has the same number N of solutions as $x^4 + y^4 \equiv -1 \pmod{p}$. By Lemma 13,

$$N = p + 1 - 6x, \quad p = x^2 + 4y^2, \quad x \equiv 1 \pmod{4}.$$

Since -2 is a quadratic non-residue of p , $h^{16} \not\equiv H^{16}$ in (181). Also $h \not\equiv 0$, $H \not\equiv 0$. If $j^{16} \equiv 1$, then $j^4 \equiv 1$. Hence the solutions fall into sets of $2 \times 4 \times 4$, so that N is a multiple of 32. There remains the case $N \leq 224$. If $N \equiv 1 \pmod{3}$, p would be divisible by 3.

If $N = 0$, $p = 5$ or 29 by III of § 37. If $N = 96$, $p = 53, 101$ or 173 by IV of § 39. Write $v = \frac{1}{2}(x - 3) = \text{odd}$. If $N = 32$, $10 = v^2 + y^2$, $p = 13, 37, 61$. If $N = 128$, $34 = v^2 + y^2$, $p = 109$ or 181. If $N = 192$, $50 = v^2 + y^2$, $v^2 = 1, 25, 49$, $p = 149, 197, 269, 293$. If $N = 224$, $58 = v^2 + y^2$, $p = 157$ or 277. For these p 's apply Lemma 9.

IV. $p \equiv 9 \pmod{16}$. By Lemma 5, (181) has the same number M of solutions as

$$(182) \quad h^8 + H^8 \equiv -1 \pmod{p}.$$

By Theorem 5, $M = 64(04)$. By (114), (115),

$$(183) \quad M = p + 1 - 18x \text{ or } M = p + 1 + 6x + 24a, \quad p = x^2 + 4y^2, \\ p = a^2 + 2b^2.$$

First, let $M > 0$. Since there exists an integer belonging to the exponent 8, $n(16) \leq 3n(14)$ unless $f(x) = x^{16} + cx^8$. As in VI, § 39, if there be a simultaneous solution of (181) and (182), then

$$H^{16} + H^8 + 1 \equiv 0, \quad H^{24} \equiv 1, \quad H^8 \not\equiv -1 \pmod{p}, \quad p \equiv 25 \pmod{48}.$$

Conversely, there are then exactly $16 \times 8 = 128$ common solutions. Hence there remains only the case $M = 128$. Then in (183₁), $p = 18x + 127$, whence 52 is the sum of the squares of $\frac{1}{2}(x - 9)$ and y , viz., 36 and 16, or vice versa. Thus $p = 73$ or $433 \equiv 1 \pmod{16}$. Next, if $M = 128$ in (183₂), then

$$p + 1 - 30\sqrt{p} < 128, \quad p < 1156.$$

But for $p = 73$ or $p < 1156$, Lemma 9 applies.

Second, let $M = 0$. Evidently $p < 1156$.

THEOREM 21. If $p > 7$, $n(16, p) \leq 5832$.

42. Case $k = 18$. Let $p \equiv -1 \pmod{3}$. By Lemma 6, with $q = 9$,

$$(184) \quad h^{18} + H^{18} \equiv -1 \pmod{p}$$

has exactly $p+1$ solutions. By Lemma 8, if $p+1 > 288$, $n(18, p) \leq 3n(16, p)$. Apply Lemma 9.

For $p > 2$, there remains the case $p \equiv 1 \pmod{6}$. If $p \equiv 1 \pmod{18}$, apply Lemma 12. Henceforth, let $p \equiv 7$ or $13 \pmod{18}$. By Lemma 5, (184) has the same number N of solutions as $x^6 + y^6 \equiv -1 \pmod{p}$. By Lemma 8, there remains the case $N \leq 18 \cdot 16 = 288$.

In the respective cases below Lemma 10,

$$\begin{aligned} p+1 &\leq -16A + 288 < 16\sqrt{p} + 288, & p < 729, \\ p+1 &\leq 288 - 2X \leq 288 + 2\sqrt{37p}, & p \leq 580. \end{aligned}$$

Apply Lemma 9. Hence we have

THEOREM 22. If $p \geq 7$, $n(18, p) \leq 17496$.

43. *Case $k=20$.* When $p \not\equiv 1 \pmod{5}$, $p \equiv -1$ or $+1 \pmod{4}$, we find by Lemma 6 or Lemma 13 the number N of solutions of $h^{20} + H^{20} \equiv -1 \pmod{p}$, and proceed as usual. If $p \equiv 1 \pmod{20}$, apply Lemma 12. There remains only the case $p \equiv 11 \pmod{20}$; since N is not known, we resort to the rough Theorem 15 and Theorem 14 and obtain

THEOREM 23. $n(20) \leq 11 n(19) \leq 22 n(18)$.

44. For $k=22$, we employ § 38 with $q=11$. It remains to treat $p \pm 1 \leq 440$; apply Lemma 9.

THEOREM 24. $n(22, p) \leq 3 n(20, p)$.

45. It remains to treat primes p which divide k . Let p^t be the highest power of p which divides k . Write

$$P = p^{t+1} \text{ if } p > 2, \quad P = p^{t+2} \text{ if } p = 2.$$

We seek N such that every integer is congruent modulo P to a sum of N values of any polynomial in x not all of whose values are multiples of p . By Lemma 9, $N \leq P-1$. Hence if $k=3$, $N \leq 8$; if $k=4$, $N \leq 15$; if $k=5$, $N \leq 24$; if $k=6$, $N \leq 8$. For these, $N \leq s$ in (160). For $7 \leq k \leq 26$, $N < n = n(k)$; for the n listed in § 29.

THE EQUIVALENCE OF NON-SINGULAR PENCILS OF HERMITIAN MATRICES IN AN ARBITRARY FIELD.

By J. WILLIAMSON.

The problem of the equivalence of two non-singular pencils of real symmetric matrices in the real field was first solved by Muth.¹ More recently Trott,² Wegner,³ Ingraham⁴ and Turnbull⁵ have solved the similar problem for two Hermitian matrices under conjunctive transformations in the complex field. The notation used by Trott was such, that he was able to discuss the Hermitian case and at the same time the real symmetric case. In this paper we show how Trott's method may be extended to the similar problem of the equivalence of two non-singular pencils of Hermitian (or symmetric) matrices with respect to a general commutative field K . Incidentally, as is often the case with a generalization, we show why the results in the case of the complex field (or real field) are comparatively simple. We prove that a necessary and sufficient condition for two such pencils to be equivalent is that;

(α) *they have the same elementary factors with respect to K ,*
and (β) *certain diagonal matrices be equivalent in over fields of K .*

In the simple cases already considered conditions (β) can all be expressed in terms of the equality of certain integers—the signatures of the respective quadratic or hermitian forms. That no such great simplification is possible in the general case is apparent from a consideration of two pairs of one rowed matrices a, b , and c, d in the rational field, where a, b, c, d , are all rational numbers and b and d are both different from zero. The pair a, b , is equivalent to the pair c, d , if, and only if, $a - \lambda b$ and $c - \lambda d$ have the same elementary

¹ P. Muth, "Über reelle Äquivalenz von Scharen reeller quadratischer Formen," *Crelle's Journal*, vol. 128 (1905), pp. 302-343.

² G. R. Trott, "On the canonical form of a non-singular pencil of Hermitian matrices," *American Journal of Mathematics*, vol. 56, no. 3 (1934), pp. 359-371. We shall refer to this paper as Trott, 1.

³ K. W. Wegner, "Equivalence of pairs of Hermitian matrices," *Bulletin of the American Mathematical Society*, vol. 40, no. 1, January (1934), Abstract 103.

⁴ M. H. Ingraham, "The singular case of the equivalence of pairs of Hermitian matrices," *Bulletin of the American Mathematical Society*, vol. 40, no. 7, July (1934), Abstract 242.

⁵ H. W. Turnbull, "Pencils of Hermitian forms," *Proceedings of the London Mathematical Society*, series 2, vol. 39 (1935), pp. 232-248.

divisors, i. e., if $a/b = c/d$ (condition α), and, if $b = k^2d$, where k is a rational number (condition β).

Section I is devoted to preliminary definitions and proofs; the main results are proved in § (2) and a short discussion of these results is given in § (3). No attempt is made to consider a similar problem for singular pencils.

1. Let K be any commutative field of characteristic zero ⁶ and let $K(i)$ be a quadratic field over K , where i is a root of the equation $x^2 - \alpha = 0$, irreducible in K . Then every element a of $K(i)$ is of the form $a = a_1 + ia_2$, where a_1 and a_2 lie in K , so that the conjugate of a is the element $\bar{a} = a_1 - ia_2$. If R is a matrix over $K(i)$, $R = R_1 + iR_2$, where R_1 and R_2 are both matrices over K , and $\bar{R} = R_1 - iR_2$. The matrix R^* is defined to be the conjugate transposed of R so that

$$R^* = \bar{R}' = R'_1 - iR'_2.$$

When R is a square matrix of order n , we may consider R as a matrix of matrices and write

$$(1) \quad R = (R_{ij}), \quad (i, j = 1, 2, \dots, t),$$

where R_{ij} is a matrix of r_i rows and r_j columns and $r_1 + r_2 + \dots + r_t = n$. If S is a second n -rowed square matrix and S is written as a matrix of matrices,

$$(2) \quad S = (S_{ij}), \quad (i, j = 1, 2, \dots, t),$$

where S_{ij} is also a matrix of r_i rows and r_j columns, we say that S and R are *similarly partitioned* or that (2) is a partition of S similar to (1). If in (1), when i is different from j , R_{ij} is the zero matrix, we call R a *diagonal block matrix* and write

$$R = [R_{11}, R_{22}, \dots, R_{tt}].$$

If D is a square matrix of order n , whose elements lie in K , the invariant factors $E_j(\lambda)$ of $D - \lambda E$ are polynomials over K . We call ⁷ the powers of the distinct irreducible factors of $E_j(\lambda)$ the elementary factors (with respect to K) of $D - \lambda E$. Let the elementary factors of $D - \lambda E$ be

$$(3) \quad [p_i(\lambda)]^{\eta_{ij}}, \\ (i = 1, 2, \dots, t; j = 1, 2, \dots, k_i; \eta_{ij} \geq \eta_{is} \geq 1, \text{ if } j < s),$$

⁶ It is not essential for this discussion that the characteristic p of K be zero. On the other hand p cannot be arbitrary. We, however, restrict ourselves to the case $p = 0$ for the sake of simplicity.

⁷ Cf. Neal McCoy, "On the rational canonical form of a function of a matrix," *American Journal of Mathematics*, this volume, p. 492; J. H. M. Wedderburn, *Lectures on Matrices*, pp. 123-126.

where $p_i(\lambda)$ is a polynomial over K of degree n_i , irreducible in K , with leading coefficient unity and such that $p_i(\lambda) \neq p_j(\lambda)$, if $i \neq j$. Then $n = \sum_{i=1}^t n_i \sum_{j=1}^{k_i} \eta_{ij}$. Further let p_i be a square matrix of order n_i , with elements in K , whose characteristic equation is $p_i(\lambda) = 0$, and let N_{ij} be the matrix

$$(4) \quad N_{ij} = \begin{pmatrix} p_i & e_i & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & p_i & e_i & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & e_i \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & p_i \end{pmatrix} \quad \begin{matrix} (i = 1, 2, \dots, t; \\ j = 1, 2, \dots, k_i), \end{matrix}$$

where e_i is the unit matrix of order n_i and N_{ij} , considered as a matrix of matrices, is of order η_{ij} . If M_i is the diagonal block matrix

$$(5) \quad M_i = [N_{i1}, N_{i2}, \dots, N_{ik_i}], \quad (i = 1, 2, \dots, t),$$

and

$$(6) \quad M = [M_1, M_2, \dots, M_t],$$

the elementary factors of $D - \lambda E$ are the same as those of $M - \lambda E$. Hence M is similar to D in K and is a canonical form of D in K .

We now define two matrices of order η_{ij} , whose elements are matrices of order n_i . These two matrices are the auxiliary unit matrix

$$(7) \quad U_{ij} = \begin{pmatrix} 0 & e_i & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & e_i & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & e_i \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \end{pmatrix}$$

and the counter unit matrix

$$(8) \quad T_{ij} = \begin{pmatrix} 0 & \cdot & \cdot & \cdot & 0 & e_i \\ 0 & \cdot & \cdot & \cdot & e_i & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ e_i & \cdot & \cdot & \cdot & 0 & 0 \end{pmatrix}.$$

The matrix N_{ij} , defined by (4), may accordingly be written in the convenient form

$$N_{ij} = p_i E_{ij} + U_{ij},$$

where $E_{ij} = T_{ij}^2$. Moreover a simple calculation shows that

$$(9) \quad U'_{ij} T_{ij} = T_{ij} U_{ij}.$$

Let q_i be a non-singular matrix over K of order n_i , satisfying the equation

$$(10) \quad q_i p_i = p'_i q_i, \quad (i = 1, 2, \dots, t).$$

It has been shown that such a matrix q_i exists and that it is necessarily symmetric.⁸ Further, if,

$$(11) \quad Q_{ij} = q_i T_{ij}, \quad (i = 1, 2, \dots, t; \quad j = 1, 2, \dots, k_i),$$

then

$$\begin{aligned} q_i T_{ij} N_{ij} &= q_i T_{ij} (p_i E_{ij} + U_{ij}), \\ &= (p'_i E_{ij} + U'_{ij}) q_i T_{ij}, \text{ by (9) and (10),} \end{aligned}$$

so that

$$(12) \quad Q_{ij} N_{ij} = N'_{ij} Q_{ij}.$$

Accordingly the matrix

$$(13) \quad Q = [Q_1, Q_2, \dots, Q_t], \quad \text{where } Q_i = [Q_{i1}, Q_{i2}, \dots, Q_{ik_i}],$$

is a non-singular symmetric matrix over K , satisfying the equation

$$(14) \quad QM = M'Q.$$

Moreover, if R is a matrix over $K(i)$, such that

$$(15) \quad RM = M'R,$$

then

$$(16) \quad R = QS,$$

where S is a matrix commutative with M . The form of S is known.⁹ In fact S is a diagonal block matrix

$$(17) \quad S = [S_1, S_2, \dots, S_t]$$

partitioned similarly to M in (6). Further, if for simplicity we write η_i, T_i, U_i, q, p , and k for $\eta_{ij}, T_{ij}, U_{ij}, q_i, p_i$ and k_i respectively and let

$$(18) \quad S_i = (S_{rs}), \quad (r, s = 1, 2, \dots, k),$$

be a partition of S_i similar to that of M_i in (5), S_{rs} is a matrix of η_r rows and η_s columns, where $\eta_r \geq \eta_s$, if $r \leq s$. Moreover, if $r \leq s$,

$$(19) \quad S_{rs} = \begin{pmatrix} G_{rs} \\ 0 \end{pmatrix}, \quad S_{sr} = (0' G_{sr}),$$

⁸ R. C. Trott, *Bulletin of the American Mathematical Society*, vol. 41, no. 1, part 2, January (1935), Abstract No. 95. We shall refer to this paper as Trott 2.

⁹ Trott 2.

where G_{rs} and G_{sr} are both square matrices of order $\eta_s = \eta$, while 0 denotes the zero matrix of orders $\eta_r - \eta$, η and $0'$ its transposed. More exactly,

$$(20) \quad G_{rs} = \sum_{a=0}^{\eta-1} g_{rsa} U_s^a, \quad G_{sr} = \sum_{a=0}^{\eta-1} g_{sra} U_s^a,$$

where g_{rsa} and g_{sra} are polynomials in the matrix p with coefficients in $K(i)$.

We now define the two matrices

$$(21) \quad \tilde{S}_{rs} = (0, \tilde{G}_{rs}), \quad \tilde{S}_{sr} = \begin{pmatrix} \tilde{G}_{sr} \\ 0 \end{pmatrix} \quad r \leq s,$$

so that in particular, if $\eta_r = \eta_s$,

$$(22) \quad \tilde{S}_{rs} = \tilde{S}_{rs}, \quad \eta_r = \eta_s.$$

It should be noted that \tilde{S}_{rs} is formally the transposed conjugate of S_{rs} , if p is considered as an indeterminate instead of a matrix.

It follows from (20) that

$$\begin{aligned} G_{rs}^* q T_s &= \sum_{a=0}^{\eta-1} g_{rsa}^* U_s'^a q T_s, \\ &= q T_s \sum_{a=0}^{\eta-1} \bar{g}_{rsa} U_s^a \text{ by (9) and (10),} \\ &= q T_s \tilde{G}_{rs}. \end{aligned}$$

Hence, if $r \leq s$,

$$\begin{aligned} S_{rs}^* q T_r &= (G_{rs}^* 0) q T_r, \quad (0 \text{ the zero matrix of orders } \eta_r - \eta_s, \eta_s) \\ &= (0 G_{rs}^* q T_s) = (0 q T_s \tilde{G}_{rs}) = q T_s (0 \tilde{G}_{rs}) = q T_s \tilde{S}_{rs} \text{ by (21).} \end{aligned}$$

Similarly,

$$S_{sr}^* q T_s = \begin{pmatrix} 0 \\ G_{sr}^* \end{pmatrix} q T_s = \begin{pmatrix} 0 \\ q T_s \tilde{G}_{sr} \end{pmatrix} = q T_r \begin{pmatrix} \tilde{G}_{sr} \\ 0 \end{pmatrix} = q T_r \tilde{S}_{sr}.$$

Therefore for all values of r and s

$$(23) \quad S_{rs}^* q T_r = q T_s \tilde{S}_{rs}.$$

If the matrix R , defined by (16), is such that $R = R^*$, on equating corresponding elements of the two matrices we have

$$q T_r S_{rs} = (q T_s S_{sr})^*,$$

or

$$(24) \quad q T_r S_{rs} = S_{sr}^* q T_s = q T_r (\tilde{S}_{sr}) \text{ by (23),}$$

so that

$$(25) \quad S_{rs} = \tilde{S}_{sr}.$$

In particular, if $\eta_r = \eta_s$, it follows from (22) that

$$(26) \quad S_{rs} = \bar{S}_{sr}, \quad (r, s = 1, 2, \dots, k)$$

and, if $r = s$, that

$$(27) \quad S_{rr} = \bar{S}_{rr}, \quad (r = 1, 2, \dots, k),$$

so that S_{rr} lies in K .

2. We now consider two square matrices, A and B , of order n , with elements in the field $K(i)$, of which the second, B , is non-singular. The matrices A and B are such that $A = A^*$ and $B = B^*$, so that both matrices are hermitian matrices or else, when A and B are both matrices over K , symmetric matrices over K . Moreover, if A and B are both matrices over K , in the sequel every matrix P is a matrix over K and P^* is to be interpreted as P' . Since $A = A^*$ and $B = B^*$, the invariant factors $E_j(\lambda)$ of the pencil $A - \lambda B$, which are certainly polynomials over the field $K(i)$, are unaltered by the substitution of $-i$ for i and are accordingly polynomials over K . We are therefore at liberty to talk of the elementary factors (with respect to K) of the pencil $A - \lambda B$. We let these elementary factors be the polynomials (3), so that $A - \lambda B$ has the same elementary factors as $M - \lambda E$. Since the elementary factors of $A - \lambda B$ are the same as those of $AB^{-1} - \lambda E$, the two matrices AB^{-1} and M are similar. Hence there exists a non-singular matrix P , such that

$$P^{-1}(AB - \lambda E)P = M - \lambda E,$$

or

$$(28) \quad (A - \lambda B)B^{-1}P = P(M - \lambda E).$$

In general the elements of the matrix P lie in $K(i)$, but, if A and B are both symmetric matrices over K , P is also a matrix over K . As a consequence of (28) we have

$$(29) \quad P^*B^{-1}(A - \lambda B)B^{-1}P = R(M - \lambda E),$$

where

$$(30) \quad R = P^*B^{-1}P.$$

It follows from (30), since $B^* = B$, that $R = R^*$ and from (29) that $RM = P^*B^{-1}AB^{-1}P$, so that RM is hermitian, and accordingly that

$$(31) \quad RM = M^*R^* = M'R.$$

We shall now reduce the pencil of matrices $R(M - \lambda E)$ by a *conjunctive*

transformation¹⁰ to a canonical form $G(M - \lambda E)$; that is, determine a non-singular matrix W such that

$$(32) \quad W^* R (M - \lambda E) W = G (M - \lambda E).$$

Accordingly, as a consequence of (29) and (32), the pencil $A - \lambda B$ is equivalent under a conjunctive transformation to the pencil $G(M - \lambda E)$. Moreover, it follows immediately from (32) that $G = W^* R W$ and that $W^* R M W = G M = W^* R W M$. Hence $G = G^*$ and

$$(33) \quad M W = W M,$$

so that throughout the various stages of the reduction the transforming matrices are all permutable with M .

As a consequence of (31) $R = Q S$, where Q is defined by (13) and S by (17), (18) and (19). Therefore,

$$R = [R_1, R_2, \dots, R_t],$$

is a diagonal block matrix, where

$$R_i = Q_i S_i, \quad (i = 1, 2, \dots, t),$$

and, since $M W = W M$, the matrix W is also a diagonal block matrix $[W_1, W_2, \dots, W_t]$, where W_i is of the same order as M_i . Hence we see that in reducing R we may reduce each R_i separately by transformations W_i permutable with M_i . As this is the case we temporarily drop all suffixes i and write M, R, S, q, T_j etc. for $M_i, R_i, S_i, Q_i, T_{ij}$ respectively.

We first show that without any loss of generality we may assume S_{11} to be non-singular. Since $\eta_1 \geq \eta_2 \geq \dots \geq \eta_k \geq 1$, we may suppose that $\eta_1 = \eta_2 = \dots = \eta_s > \eta_{s+1}$, $1 \leq s \leq k$. If S_{11} is singular but, for some value of $j \leq s$, S_{jj} is non-singular, by interchanging the first row and the j -th row of S and the first column and the j -th column, we move S_{jj} into the place of S_{11} . Moreover such an interchange may be accomplished by means of a conjunctive transformation permutable with M and Q .¹¹ We now assume that S_{jj} is singular for all values of j , $1 \leq j \leq s$. If s_{ij} denotes the first element, i. e., the element in the first row and first column, of the matrix S_{ij} , $|S_{jj}| = |s_{jj}|^n$ and, since s_{jj} is a polynomial in the matrix p with coefficients

¹⁰ We shall use the term conjunctive transformation to include the case of a congruent transformation; i. e., a transformation of matrix W , where W lies in K , so that $W^* = W'$.

¹¹ Turnbull and Aitken, *An Introduction to the Theory of Canonical Matrices*, p. 11.

in K (equation (27)), $s_{jj} = 0$. In particular $s_{11} = 0$, and, since, from the nature of S_{i1} (equation (19)), $s_{i1} = 0$ when $i > s$, there is at least one value j , $1 < j \leq s$, such that $s_{j1} \neq 0$, as otherwise S would be singular. After a suitable interchange of rows and columns we may therefore suppose that s_{21} is not zero. Let

$$W_1 = \left[\begin{pmatrix} E_1 & 0 \\ E_1 & E_1 \end{pmatrix}, E_2 \right] \quad \text{and} \quad W_2 = \left[\begin{pmatrix} E_1 & 0 \\ iE_1 & E_1 \end{pmatrix}, E_2 \right],$$

where E_1 and E_2 are the unit matrices of orders $n_i\eta_1$ and $n_i(\eta_3 + \eta_4 + \dots + \eta_n)$ respectively. The two matrices W_1 and W_2 are both permutable with M as are the matrices W^*_1 and W^*_2 with Q . A simple calculation shows that, if

$$W^*_1 Q S W_1 = QX \quad \text{and} \quad W^*_2 Q S W_2 = QY$$

and X and Y are partitioned similarly to M ,

$$X_{11} = S_{11} + S_{21} + S_{12} + S_{22}, \quad Y_{11} = S_{11} + i(S_{12} - S_{21}) - i^2 S_{22}.$$

The first two elements of these matrices are respectively $x_{11} = s_{12} + s_{21}$ and $y_{11} = i(s_{12} - s_{21})$, since $s_{11} = s_{22} = 0$. As $s_{21} \neq 0$, at least one of x_{11} or y_{11} is different from zero, so that at least one of X_{11} or Y_{11} is non-singular. However, as W_2 is not a matrix over K , we must still show that, if S is a matrix over K , the matrix X_{11} is non-singular. This is in fact the case; for, since $S^* = S$, by (26) $s_{21} = \bar{s}_{12}$, so that if S lies in K , $s_{21} = \bar{s}_{12} = s_{12}$ and $x_{11} = 2s_{21} \neq 0$. Hence we may assume without any loss of generality that S_{11} is non-singular.

We next show that S may be reduced to a diagonal block matrix partitioned similarly to M . Let

$$W = \begin{pmatrix} E_1 & -S_{11}^{-1}S_{12} & -S_{11}^{-1}S_{13} & \dots & -S_{11}^{-1}S_{1k} \\ 0 & E_2 & 0 & \dots & 0 \\ 0 & 0 & E_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & E_k \end{pmatrix},$$

so that W is certainly permutable with M . The element in the r -th place, $r > 1$, of the first column of W^*Q is

$$-S^*_{1r}(S^*_{11})^{-1}qT_1 = -S^*_{1r}qT_1S_{11}^{-1} = -qT_rS_{r1}S_{11}^{-1} \text{ by (24).}$$

Hence $W^*Q = QH$, where H is obtained from W^* by replacing the element in the r -th place, $r > 1$, of the first column by $-S_{r1}S_{11}^{-1}$. A simple calculation now shows that if,

$$\begin{aligned}
W^*QSW &= QD = Q(D_{rs}), & (r, s = 1, 2, \dots, k), \\
D_{11} &= S_{11}, \quad D_{r1} = D_{1r} = 0, & (r \neq 1), \\
D_{rs} &= S_{rs} - S_{r1}S_{11}^{-1}S_{1s}, & (r, s = 2, 3, \dots, k).
\end{aligned}$$

We have therefore shown that there exists a non-singular conjunctive transformation permutable with M , which reduces $R = QS$ to the form $Q[S_{11}, H]$, where H is a square matrix of order $n_1(\eta_2 + \eta_3 + \dots + \eta_k)$ and is commutative with $[N_2, N_3, \dots, N_k]$. Accordingly H is of exactly the same type as S and we may repeat our previous argument with S replaced by H . Hence in $k - 1$ steps we deduce the existence of a non-singular matrix V , permutable with M , such that

$$(34) \quad V^*QSV = Q[G_1, G_2, \dots, G_k],$$

where G_j is of the same order as N_j , ($j = 1, 2, \dots, k$). Moreover G_j is permutable with N_j and $Q_j G_j$ is hermitian. Hence by (27) G_j is a matrix over K and $Q_j G_j$ is symmetric.

We now show that it is possible to reduce G_j , ($j = 1, 2, \dots, k$), to a diagonal block matrix. For simplicity we write $\eta = \eta_j$, so that

$$G_j = \sum_{a=0}^{\eta-1} g_a U_j^a \quad (\text{formulas (19) and (20)}),$$

where g_a is a polynomial in p with coefficients in K . Since G_j is non-singular, g_0 is non-singular and is accordingly different from zero. If $g_1 = g_2 = \dots = g_c = 0$ and $c = \eta - 1$, G is a diagonal block matrix. If $c < \eta - 1$ and $g_{c+1} \neq 0$, we consider the matrix

$$W = E_j - w U_j^{c+1}, \quad \text{where } w = g_{c+1}/2g_0,$$

so that

$$\begin{aligned}
G_j W^2 &= G_j (E_j - 2w U_j^{c+1} + w^2 U_j^{2c+2}), \\
&= g_0 E_j + h_{c+1} U_j^{c+1} + \sum_{a=c+2}^{\eta-1} h_a U_j^a = H_j,
\end{aligned}$$

where $h_{c+1} = g_{c+1} - 2gw = 0$ and h_a is a polynomial in p , when $a \geq c + 2$. But

$$\begin{aligned}
W^* Q_j G_j W &= W^* q T_j G_j W = q T_j W G_j W \text{ by (9) and (10),} \\
&= q T_j G_j W^2 = q T_j H_j.
\end{aligned}$$

Hence the matrix W reduces $Q_j G_j$ to $Q_j H_j$, where H_j is of the same form as G_j except that the coefficient of U_j^{c+1} is now zero. If the coefficient of U_j^{c+2} in H_j is different from zero, we may repeat our argument with G_j replaced

by H_j . Accordingly in at most $\eta - 1$ such steps we can reduce G_j to the diagonal form $g_0 E_j$. Let us therefore suppose that G_j is already in diagonal form, so that, after an obvious change of notation,

$$(35) \quad G_j = g_j E_j, \quad (j = 1, 2, \dots, k),$$

and $g_j = g_j(p)$ is a polynomial in p with coefficients in K . Equation (34) therefore becomes,

$$(36) \quad V^* Q S V = Q G = Q [g_1 E_1, g_2 E_2, \dots, g_k E_k],$$

and we call the matrix on the right of this last equation a *canonical form* for the matrix $Q S$.

It is apparent that this canonical form may not be unique. Suppose therefore that there exists a non-singular matrix Y , permutable with M , such that

$$(37) \quad Y^* Q S Y = Q F = Q [f_1 E_1, f_2 E_2, \dots, f_k E_k],$$

where $f_j = f_j(p)$ is a polynomial in p with coefficients in K . Then, if $W = V^{-1} Y$, W is permutable with M and, as a consequence of (36) and (37),

$$(38) \quad W^* Q G W = Q F.$$

If $W = (W_{rs})$, ($r, s = 1, 2, \dots, k$), is a partition of W similar to that of M , W_{rs} and W_{sr} are of the same forms as S_{rs} and S_{sr} in (19). We define \bar{W}_{rs} in an analogous manner to \bar{S}_{rs} in (21), so that in particular, if $\eta_r = \eta_s$, $\bar{W}_{rs} = \bar{W}_{sr}$. The matrix equation (38) may now be written in the form

$$\sum_{a=1}^k W_{ar}^* q T_a g_a W_{as} = \delta_{rs} q T_r f_r E_r, \quad (r, s = 1, 2, \dots, k; \delta_{rs} \text{ the Kronecker } \delta).$$

Hence by (23)

$$q T_r \sum_{a=1}^k \bar{W}_{ar} g_a W_{as} = \delta_{rs} q T_r f_r E_r,$$

or, on dividing by the non-singular matrix $q T_r$,

$$(39) \quad \sum_{a=1}^k \bar{W}_{ar} g_a W_{as} = \delta_{rs} f_r E_r.$$

If w_{ij} is the first element of the matrix W_{ij} and \bar{w}_{ij} the first element of the matrix \bar{W}_{ij} , it follows from the nature of the matrices W_{ij} and \bar{W}_{ij} (cf. equations (19) and (21)), that the first element of the matrix $\bar{W}_{ar} g_a W_{as}$ is $\bar{w}_{ar} g_a w_{as}$. Accordingly by equating the first elements of each component matrix in (39), we have

$$(40) \quad \sum_{a=1}^k \bar{w}_{ar} g_a w_{as} = \delta_{rs} f_r.$$

But by (19) and (21),

$$w_{as} = 0, \text{ if } \eta_a < \eta_s; \bar{w}_{ar} = 0, \text{ if } \eta_a > \eta_r; \bar{w}_{ar} = \bar{w}_{ar}, \text{ if } \eta_a = \eta_r.$$

Hence, if $\eta_{c-1} > \eta_c = \eta_{c+1} = \dots = \eta_d > \eta_{d+1}$ and $c \leq s \leq d$, $c \leq r \leq d$, $\bar{w}_{ar} g_a w_{as} = 0$, when $\alpha < c$ or $\alpha > d$. Accordingly (40) becomes

$$(41) \quad \sum_{a=c}^d \bar{w}_{ar} g_a w_{as} = \delta_{rs} f_r.$$

Let D be the matrix (d_{ij}) , $(i, j = 1, 2, \dots, d - c + 1)$, where $d_{ij} = w_{c+i-1, c+j-1}$. Then it is a consequence of the form of W that, since W is non-singular, D is non-singular, for, after a proper interchange of rows and columns, it can be shown that $|D|$ is a factor of $|W|$. Each element d_{ij} of D is a polynomial $\bar{d}_{ij}(p)$ in the matrix p with coefficients in $K(i)$ and therefore (41) may be written in the form of a matrix equation

$$(42) \quad \bar{D}[g_c, g_{c+1}, \dots, g_d] D = [f_c, f_{c+1}, \dots, f_d],$$

where $\bar{D} = (\bar{d}_{ij}) = (\bar{d}_{ji})$, $i, j = 1, 2, \dots, d - c + 1$; cf. equation (21). It is important to notice that \bar{D} is not the same as D^* , since $D^* = (d^*_{ij})$, where $d^*_{ij} = \bar{d}_{ji}(p')$. Let x be an indeterminate and let $D(x)$ denote the matrix whose typical element is $d_{ij}(x)$. Then, if

$$(43) \quad |D(x)| = \rho(x) + i\sigma(x), \rho(x), \sigma(x) \text{ polynomials with coefficients in } K,$$

$$(44) \quad |D| = |\rho(p) + i\sigma(p)|.^{12}$$

Similarly $|\bar{D}| = |\rho(p) - i\sigma(p)|$, so that

$$(45) \quad |D| |\bar{D}| = |(\rho(p))^2 - i^2(\sigma(p))^2| = |\mu(p)|,$$

where $\mu(x)$ is a polynomial in x with coefficients in K . Since D , and similarly, \bar{D} , are both non-singular, $D\bar{D}$ is non-singular, so that by (45), $\mu(p)$ is non-singular. Hence, since $\mu(p)$ is a polynomial over K ,

$$(46) \quad \mu(p) \neq 0,$$

is a necessary and sufficient condition that D and \bar{D} both be non-singular.

If θ is a root of the irreducible equation $p(x) = 0$, the field $K(\theta)$ is simply isomorphic with the field formed by all polynomials in p with coefficients in K . Consequently it follows from (45) and (46) that

¹² J. Williamson, "The latent roots of a matrix of special type," *Bulletin of the American Mathematical Society*, vol. 37 (August, 1931), p. 587.

$$(47) \quad |D(\theta)| \mid \widetilde{D}(\theta) = \mu(\theta) \neq 0$$

and accordingly that both matrices $D(\theta)$ and $\widetilde{D}(\theta)$ are non-singular. Since the elements of $D(\theta)$ are no longer matrices, $\widetilde{D}(\theta) = D^*(\theta)$, and we therefore have, as a consequence of (42),

$$(48) \quad D^*(\theta)[g_c(\theta), \dots, g_d(\theta)]D(\theta) = [f_c(\theta), \dots, f_d(\theta)],$$

where $D(\theta)$ and $D^*(\theta)$ are both non-singular. In other words the two matrices $[g_c(\theta), \dots, g_d(\theta)]$ and $[f_c(\theta), \dots, f_d(\theta)]$ are conjunctively equivalent. Conversely, if (48) is true and both $D(\theta)$ and $D^*(\theta)$ are non-singular,¹³ $\mu(\theta) \neq 0$ by (47) and accordingly (46) is satisfied, so that (42) is true, where D is non-singular. Hence not only does (42) imply (48) but also (48) implies (42).

Before summing up and stating our results in the form of a theorem it will prove convenient to alter our notation slightly. Accordingly we relabel the integers η_i in the following manner;

$$(49) \quad \eta_1 = \eta_2 = \dots = \eta_{s_1} = \xi_1 > \eta_{s_1+1} = \eta_{s_1+2} = \dots = \eta_{s_1+s_2} = \xi_2 > \eta_{s_1+s_2+1} \\ = \dots = \xi_{r-1} > \eta_{s_1+s_2+\dots+s_{r-1}+1} = \dots = \eta_{s_1+s_2+\dots+s_r} = \xi_r,$$

where $s_1 + s_2 + \dots + s_r = k$, and write

$$(50) \quad \begin{aligned} I_j &= [E_c, E_{c+1}, \dots, E_d], & L_j &= [N_c, N_{c+1}, \dots, N_d], \\ \gamma_j &= [g_c, g_{c+1}, \dots, g_d], & \phi_j &= [f_c, f_{c+1}, \dots, f_d], \\ \Gamma_j &= [g_c E_c, g_{c+1} E_{c+1}, \dots, g_d E_d], & \Phi_j &= [f_c E_c, f_{c+1} E_{c+1}, \dots, f_d E_d], \\ Q_j^{(1)} &= q[T_c, T_{c+1}, \dots, T_d], & & (j=1, 2, \dots, r), \end{aligned}$$

where $c = s_1 + s_2 + \dots + s_{j-1} + 1$, $d = s_1 + s_2 + \dots + s_j$. Using this notation we may express our last result in the form of a lemma;

LEMMA I. *If the two canonical forms QG and QF of equations (37) and (38) respectively, are equivalent, there exist $2r$ non-singular matrices $D_j(\theta)$, $D_j^*(\theta)$ with elements in $K(\theta, i)$ such that,*

$$(51) \quad D_j^*(\theta)\gamma_j(\theta)D_j(\theta) = \phi_j(\theta), \quad (j=1, 2, \dots, r);$$

i. e., the matrices $\gamma_j(\theta)$, $\phi_j(\theta)$, ($j=1, 2, \dots, r$), are equivalent under a non-singular conjunctive transformation in the field $K(\theta, i)$.

The converse of this lemma is also true. In fact (51) implies that

¹³ If i lies in the field $K(\theta)$, $|D(\theta)| \neq 0$ does not imply $|D^*(\theta)| \neq 0$.

$\bar{D}_j \gamma_j D_j = \phi_j$, ($j = 1, 2, \dots, r$), where $|D_j| \neq 0$. If W_j is the matrix obtained from D_j by replacing each element d of D_j by $E_{s_1+s_2+\dots+s_j}$, it immediately follows that

$$\bar{W}_j \Gamma_j W_j = \Phi_j, \quad (j = 1, 2, \dots, r),$$

and since $Q_j^{(1)} \bar{W}_j = W_j^* Q_j^{(1)}$, that

$$W_j^* Q_j^{(1)} \Gamma_j W_j = Q_j^{(1)} \Phi_j.$$

Hence, if $W = [W_1, W_2, \dots, W_r]$, W is non-singular and $W^* Q G W = Q F$, so that the two normal forms $Q G$ and $Q F$ are equivalent.

In stating the theorem given below we use a notation conforming with that explained in (49) and (50); the matrices defined in (50) are associated with a particular polynomial $p_i(\lambda)$ and for convenience in writing we dropped the suffix i but now we find it necessary to replace it. We have proved the theorem:

THEOREM I. *Let A and B be two matrices, of which the second B is non-singular, with elements in $K(i)$ and let $A = A^*$ and $B = B^*$. If the elementary factors of $A - \lambda B$ are the polynomials $[p_i(\lambda)]^{s_{ij}}$ of (3), then a canonical form for the pencil $A - \lambda B$ under a non-singular conjunctive transformation is the diagonal block matrix*

$$(52) \quad Q G (M - \lambda E) = Q [\Gamma_{ij} (L_{ij} - \lambda I_{ij})], \quad (i = 1, 2, \dots, t; j = 1, 2, \dots, r_i),$$

where Q is defined by (13) while L_{ij} , Γ_{ij} and I_{ij} are defined by (50). Two canonical forms $Q G (M - \lambda E)$ and $Q F (M - \lambda E)$, where $F = [\Phi_{ij}]$, ($i = 1, 2, \dots, t; j = 1, 2, \dots, r_i$), are equivalent, if and only if the diagonal matrices $\gamma_{ij}(\theta_i)$ and $\phi_{ij}(\theta_i)$ are equivalent under a conjunctive transformation in the field $K(\theta_i, i)$, ($i = 1, 2, \dots, t; j = 1, 2, \dots, r_i$).

Thus, if $[p_i(\lambda)]^{s_{ij}}$ occurs exactly s_{ij} times among the elementary factors of $A - \lambda B$, in a canonical form (52), there is associated with this elementary factor a diagonal matrix $\gamma_{ij}(\theta_i)$ of order s_{ij} , whose elements lie in the field $K(\theta_i)$, where θ_i is a root of the irreducible equation $p_i(\lambda) = 0$. The matrix $\gamma_{ij}(\theta_i)$ is determined apart from a conjunctive transformation.

Throughout we have used conjunctive transformation to include the case of congruent transformation. We therefore see that, if A and B are symmetric matrices over K , Theorem I is true when 'conjunctive' is replaced by 'congruent' and $K(i)$ is replaced by K .

We now state two corollaries of Theorem 1.

COROLLARY I. We may determine the matrices γ_{ij} of a canonical form (52) in such a way that no element of γ_{ij} contains a factor r^2 , where r is a polynomial in the matrix p with coefficients in K .

For, if $\phi_{ij} = \rho_{ij}^2 \gamma_{ij}$, and ρ_{ij} is a diagonal matrix whose elements are polynomials in p_i with coefficients in K , $\rho_{ij}^* \gamma_{ij} \rho_{ij} = \phi_{ij}$, so that ϕ_{ij} is equivalent to γ_{ij} .

COROLLARY II. Two pairs of hermitian matrices A, B and C, D , with elements in $K(i)$, the second of each pair being non-singular, are equivalent under a non-singular conjunctive transformation in $K(i)$, if, and only if, the two pencils $A - \lambda B$ and $C - \lambda D$ have the same elementary factors and, if the matrices $\gamma_{ij}(\theta_i)$, associated with each distinct elementary factor, are conjunctively equivalent in $K(i, \theta_i)$.

COROLLARY III. Corollary II remains true if hermitian is replaced by symmetric, conjunctive by congruent and $K(i)$ by K .

We may use Theorem 1 to determine a canonical form for any non-singular pencil of matrices $A - \lambda B$ with elements in $K(i)$. For, if B is singular but $|A - \lambda B| \not\equiv 0$, we may determine a new basis for the pencil, A_1 and B_1 , where B_1 is non-singular and $A - \lambda B = A_1 - \rho B_1$.¹⁴ We apply Theorem I to the pencil $A_1 - \rho B_1$ and thus determine a canonical form for the non-singular pencil $A - \lambda B$.

3. *Ordinary hermitian matrices and real symmetric matrices.* If K is the field of all real numbers and $K(i)$ the complex number field, the polynomials $p_i(\lambda)$ of (3) are either quadratic or linear. If $p_i(\lambda)$ is quadratic, $K(\theta_i) = K(i)$ and hence, if $\gamma_{ij}(\theta_i)$ is one of the matrices associated with $p_i(\lambda)$, $\gamma_{ij}(\theta_i)$ is a diagonal matrix, whose elements are complex numbers. Let $\gamma_{ij}(\theta_i) = [g_r]$ and let $W = [w_r]$, where $w_r = g_r^{1/2}$, if $g_r \neq 0$, and $w_r = 1$, if $g_r = 0$. Then the matrix W is a non-singular matrix with elements in $K(\theta_i)$, as is the matrix W^{-1} . But $(W^{-1})' \gamma_{ij}(\theta_i) W^{-1}$ is the identity matrix. Hence each matrix $\gamma_{ij}(\theta_i)$ associated with $p_i(\lambda)$ may be reduced to the identity matrix of the corresponding order. If $p_i(\lambda) = \lambda^2 - 2a_i\lambda + a_i^2 + b_i^2$, we may choose for p_i the matrix $\begin{pmatrix} a_i & -b_i \\ b_i & a_i \end{pmatrix}$ and for q_i the matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. If, however, $p_i(\lambda)$ is linear, $K(\theta_i) = K$, the field of all real numbers and by Corollary 2 each associated $\gamma_{ij}(\theta_i)$ may be reduced to a diagonal matrix with elements, which are either $+1$ or -1 . If $p_i(\lambda) = \lambda - \lambda_i$, $p = \lambda_i$ and

¹⁴ Cf. Turnbull and Aitken, *op. cit.*, p. 117 sq.; Trott 1, p. 370.

$q = 1$. The normal form (52) therefore coincides with the normal form given by Trott 1, page 368, formula (11). In this particular case, however, the condition of Lemma 1 is greatly simplified, for two diagonal matrices with real coefficients are equivalent under a conjunctive or congruent transformation, if, and only if, they have the same signature. Thus Trott's condition (15) merely expresses the fact that $\gamma_{ij}(\theta_i)$ is conjunctively equivalent to $\phi_{ij}(\theta_i)$.

In the general case no such simplification of the conditions in Lemma 1 is possible. The conditions for the equivalence of two quadratic or hermitian forms have been determined but are very complicated.¹⁵ We however state necessary and sufficient conditions in the two simplest cases (a) $s_{ij} = 1$, (b) $s_{ij} = 2$. These conditions are due to Dickson.

(a) If $\gamma_{ij}(\theta_i)$ is of order one, $\gamma_{ij}(\theta_i)$ is an element of $K(\theta_i)$. Then $\gamma_{ij}(\theta_i)$ is equivalent to $\phi_{ij}(\theta_i)$ if, and only if, there exists an element f of $K(i, \theta_i)$, such that

$$\phi_{ij}(\theta_i) = f\bar{f}\gamma_{ij}(\theta_i).$$

(b) If the matrices $\gamma_{ij}(\theta_i)$ and $\phi_{ij}(\theta_i)$ are both of order two, they may be represented as $\begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix}$ and $\begin{pmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{pmatrix}$ respectively. Then $\gamma_{ij}(\theta_i)$ is equivalent to $\phi_{ij}(\theta_i)$, if, and only if, there exist elements f, g and h of $K(\theta, i)$, such that $\phi_1 = \gamma_1 f\bar{f} + \gamma_2 g\bar{g}$ and $\phi_1\phi_2 = h\bar{h}\gamma_1\gamma_2$. In the symmetric case the elements f, g, h lie in K and $f = \bar{f}, g = \bar{g}, h = \bar{h}$.

4. We conclude our discussion by giving an explicit form for the matrices, p_i and q_i , which occur in the canonical form (52). The matrix $p = p_i$ is a matrix of order $n = n_i$, whose characteristic equation is the irreducible equation $p_i(\lambda) = p(\lambda) = 0$ of degree n . If

$$p(\lambda) = \lambda^n - a_n\lambda^{n-1} - a_{n-1}\lambda^{n-2} \cdots - a_2\lambda - a_1,$$

we may take for the matrix p the companion matrix of $p(\lambda)$,

$$p = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_1 \\ 1 & 0 & \cdots & 0 & a_2 \\ 0 & 1 & \cdots & 0 & a_3 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & a_n \end{pmatrix},$$

¹⁵ L. E. Dickson, "On quadratic, bilinear, and Hermitian forms," *Transactions of the American Mathematical Society*, vol. 7 (1906), pp. 275-292; "On quadratic forms in a general field," *Bulletin of the American Mathematical Society*, vol. 14 (1907-8), pp. 108-115; H. Hasse, "Symmetrische Matrizen in Körper der rationalen zahlen," *Crelle*, vol. 153, pp. 12-43.

and for $q = q_1$ the matrix

$$(53) \quad q = \begin{pmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 & b_2 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & b_2 & b_3 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & b_2 & b_3 & b_4 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & b_2 & b_3 & \cdot & \cdot & \cdot & b_{n-2} & b_{n-1} & b_n \end{pmatrix},$$

where $b_2 = a_1$ and $b_{i+1} = b_2 a_{n+2-i} + b_3 a_{n+3-i} + \cdots + b_i a_n$, ($i = 2, 3, \cdots, n-1$). It is obvious that q is non-singular, for p is non-singular and hence $b_2 = a_1 \neq 0$. Moreover it is easily verified that $qp = p'q$. The matrix q is not uniquely determined by the matrix p , but the matrix q in (53) is as simple in form as, if not simpler than, any other matrix q_1 satisfying the equation $p'q_1 = q_1p$.

THE JOHNS HOPKINS UNIVERSITY.

ON THE RATIONAL CANONICAL FORM OF A FUNCTION OF A MATRIX.

By NEAL H. MCCOY.

Let A be a matrix of order n with elements in the complex number field, and $\phi(A)$ a given rational integral function of A . In 1906, Kreis¹ gave a method of determining the elementary divisors of $\phi(A)$ from those of A . In recent years the same problem has been discussed by Krishnamurthy,² Turnbull and Aitken,³ Rutherford⁴ and Amante.⁵ It is sufficient to consider the case in which A has a single elementary divisor $(\lambda - a)^n$, as the general case easily reduces to this one. The principal result of these writers may then be stated in the following way. Expand $\phi(\lambda)$ in powers of $\lambda - a$,

$$\phi(\lambda) = a_0 + a_1(\lambda - a) + a_2(\lambda - a)^2 + \cdots$$

Suppose the i -th number of the sequence $a_1, a_2, \dots, a_{n-1}, 1$, is the first which is not zero. Define positive integers k and l by the relations,

$$n = (k - 1)i + l, \quad k \geq 1, \quad 1 \leq l \leq i.$$

Then $\phi(A)$ has the elementary divisors $(\lambda - a_0)^k$ taken l times, and $(\lambda - a_0)^{k-1}$ taken $i - l$ times.

So far as the writer is aware, no solution has been given of the problem corresponding to this one, for the case in which the elements of A and all operations are restricted to an arbitrary domain of rationality. In this more general problem one does not have the use of the comparatively simple Jordan normal form of a matrix, and a different method of attack must therefore be used. It is the purpose of the present paper to present a solution of this problem.

In § 4, we shall also give a brief account of an application of the main result to the solution of certain matrix equations.

1. The rational canonical form. Let K denote a given field. Unless

¹ H. Kreis, *Contribution à la théorie des systèmes linéaires*, Zürich, 1906.

² Rao S. Krishnamurthy, "Invariant-factors of a certain class of linear substitutions," *Journal of the Indian Mathematical Society*, vol. 19 (1932), pp. 233-240.

³ H. W. Turnbull and A. C. Aitken, *Canonical Matrices*, Glasgow, 1932, pp. 75-76.

⁴ D. E. Rutherford, "On the canonical form of a rational integral function of a matrix," *Proceedings of the Edinburgh Mathematical Society* II, vol. 3 (1932), pp. 135-143.

⁵ S. Amante, "Sulle riduzione a forma canonica di una classe speciale di matrici," *Atti della Reale Accademia Nazionale dei Lincei*, Rendiconti VI, vol. 17 (1933), pp. 31-36 and pp. 431-436.

otherwise stated, it will be assumed henceforth that all matrices and vectors have coordinates in K , and all polynomials have coefficients in K . If a polynomial is irreducible relative to the field K , we shall simply say that it is irreducible.

Let $f(\lambda) = \lambda^p - a_1\lambda^{p-1} - \cdots - a_p$ be a given polynomial, and form the matrix,

$$B = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_p & a_{p-1} & a_{p-2} & \cdots & a_1 \end{pmatrix}.$$

This matrix may be called the *companion matrix* of the function $f(\lambda)$ or of the equation $f(\lambda) = 0$.⁶ The minimum function of B is then $|\lambda - B| = f(\lambda)$.

Let now A be a given matrix of order n , and $E_i(\lambda)$, ($i = 1, 2, \cdots, r$), the non-constant invariant factors of $\lambda - A$. Perhaps the most common rational canonical form of A (with respect to similarity transformations) is a matrix A_1 , which is the direct sum⁷ of the companion matrices of the $E_i(\lambda)$. However, it will be convenient for our purpose to use a somewhat different rational canonical form, which will now be described.

If we factor the $E_i(\lambda)$ into powers of distinct, irreducible polynomials $p_k(\lambda)$, each of which has leading coefficient unity, say

$$E_i(\lambda) = [p_1(\lambda)]^{n_{i1}} [p_2(\lambda)]^{n_{i2}} \cdots [p_l(\lambda)]^{n_{il}} \quad (i = 1, 2, \cdots, r),$$

then such of the factors $[p_k(\lambda)]^{n_{ik}}$ as are not mere constants may be called the *elementary divisors* of A . We can then choose as a canonical form of A , a matrix A_2 , which is the direct sum of the companion matrices of the elementary divisors of A .⁸ This is the canonical form used throughout this paper. The advantage of this form over the other lies in the fact that if $A = C \dot{+} D$, the canonical form of A is the direct sum of the canonical forms of C and D , and the elementary divisors of A are the elementary divisors of C , together with those of D .

⁶ See C. C. MacDuffee, *The Theory of Matrices*, Berlin, 1933, p. 20.

⁷ If $A = \begin{pmatrix} C & 0 \\ 0 & D \end{pmatrix}$, where C and D are square matrices, then A is called the *direct sum* of C and D , and we write, $A = C \dot{+} D$.

⁸ W. Krull, "Theorie und Anwendung der verallgemeinerten Abelschen Gruppen," *Sitzungsberichte Heidelberger Akademie der Wissenschaften*, 1926, pp 25-28; B. L. van der Waerden, *Moderne Algebra*, vol. 2, Berlin, 1931, p. 137. For a somewhat different canonical form, but one which also uses the notion of elementary divisors, see J. H. M. Wedderburn, "Note on matrices in a given field," *Annals of Mathematics*, vol. 27 (1926), pp. 245-248.

We shall now establish two lemmas^{*} which will be useful also in a later section of the paper, and then apply them to show how to find a non-singular matrix which transforms A_1 into A_2 .

LEMMA 1. Let V_j be a given row vector of dimension n , X an arbitrary column vector of dimension n , and B a given square matrix of order n . Denote by R_j the matrix of e_j rows and n columns, whose rows are respectively the vectors

$$V_j, V_j B, V_j B^2, \dots, V_j B^{e_j-1}.$$

If now $h_j(\lambda) = \lambda^{e_j} - b_1 \lambda^{e_j-1} - \dots - b_{e_j}$, is a polynomial such that $V_j h_j(B) = 0$, and we set $\xi_j = R_j X$, $Y = BX$, $\eta_j = R_j Y$, then it follows that $\eta_j = Q_j \xi_j$, where Q_j is the companion matrix of $h_j(\lambda)$.

By definition, we see that $\xi_j = \{\xi_{j1}, \xi_{j2}, \dots, \xi_{je_j}\}$ and $\eta_j = \{\eta_{j1}, \eta_{j2}, \dots, \eta_{je_j}\}$ are column vectors of dimension e_j . The lemma follows at once from the following calculation:

$$\begin{aligned} V_j X &= \xi_{j1}, \\ \eta_{j1} = V_j B X &= \xi_{j2}, \\ \eta_{j2} = V_j B^2 X &= \xi_{j3}, \\ &\vdots \\ \eta_{j, e_j-1} = V_j B^{e_j-1} X &= \xi_{je_j}, \\ \eta_{j, e_j} = V_j B^{e_j} X &= b_1 \xi_{j, e_j-1} + b_2 \xi_{j, e_j-2} + \dots + b_{e_j} \xi_{j1}. \end{aligned}$$

LEMMA 2. Let e_j ($j = 1, 2, \dots, q$) be positive integers whose sum is n , and for each j suppose V_j , R_j , ξ_j , η_j , $h_j(\lambda)$, Q_j , satisfy the conditions of Lemma 1. If we set

$$R = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_q \end{pmatrix}, \quad \xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_q \end{pmatrix}, \quad \eta = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_q \end{pmatrix},$$

then $\xi = RX$, $Y = BX$, $\eta = RY$, and $\eta = Q\xi$, where $Q = Q_1 + Q_2 + \dots + Q_q$. Further, if R is non-singular, then $Q = RBR^{-1}$.

The first part follows almost immediately from the preceding lemma. We then find that $RBX = QRX$. But since X is entirely arbitrary, we must have $RB = QR$. Hence if R is non-singular, $Q = RBR^{-1}$.

^{*} I am indebted to a referee for suggesting the introduction of these lemmas. Their use has considerably improved the proof of Theorem 1.

Let $U = (u_1, u_2, \dots, u_n)$ be any row vector. Then there exists a unique polynomial $g(\lambda)$ of minimum degree and with leading coefficient unity, such that $Ug(A) = 0$.¹⁰ This polynomial $g(\lambda)$ is called the *R. C. F. (Reduced Characteristic Function)* of A relative to U , and its degree may be called the *grade* of U (relative to A). A fundamental property of the R. C. F. of A relative to U is that, if $h(\lambda)$ is a polynomial such that $Uh(A) = 0$, then $h(\lambda)$ is divisible by $g(\lambda)$.

We now return to the problem of transforming the matrix A_1 into A_2 . Since A_1 is the direct sum of the companion matrices of the invariant factors of $\lambda - A$, we may assume for our purpose, that $\lambda - A$ has a single invariant factor $E(\lambda)$ of degree n . If $E(\lambda)$ is a power of an irreducible polynomial, then clearly $A_1 = A_2$. Hence suppose $E(\lambda) = \phi(\lambda)\psi(\lambda)$, where $\phi(\lambda)$ and $\psi(\lambda)$ are relatively prime, and have leading coefficients equal to unity. Let the degrees of $\phi(\lambda)$ and $\psi(\lambda)$ be respectively n_1 and n_2 .

From the form of A_1 , it follows that the vector, $U_1 = (1, 0, \dots, 0)$, is of grade n relative to A_1 , and the R. C. F. of A_1 relative to U_1 is therefore $E(\lambda)$. We may now apply the above lemmas by placing $e_1 = n_2$, $e_2 = n_1$, $B = A_1$, $V_1 = U_1\phi(A_1)$, $V_2 = U_1\psi(A_1)$. Since $U\phi(A_1)\psi(A_1) = 0$, it follows at once that $h_1(\lambda) = \psi(\lambda)$, $h_2(\lambda) = \phi(\lambda)$. If now we assume for the moment that R is non-singular, we see by means of Lemma 2 that $RA_1R^{-1} = Q$, where Q is the direct sum of the companion matrices of $\phi(\lambda)$ and of $\psi(\lambda)$. If either $\phi(\lambda)$ or $\psi(\lambda)$ can be expressed as a product of relatively prime factors, the process can be continued, and so on until the form A_2 is reached.

We now show that R is non-singular. For suppose there exists a relation

$$U_1 \left[\sum_{i=0}^{n_2-1} c_i \phi(A_1) A_1^i + \sum_{i=0}^{n_1-1} d_i \psi(A_1) A_1^i \right] = 0.$$

Since $E(\lambda)$ is the R. C. F. of U_1 relative to A_1 , this implies that the polynomial

$$F(\lambda) \equiv \phi(\lambda) \sum_{i=0}^{n_2-1} c_i \lambda^i + \psi(\lambda) \sum_{i=0}^{n_1-1} d_i \lambda^i,$$

is divisible by $E(\lambda)$, and being of degree at most $n - 1$, must therefore vanish identically. From the fact that $\phi(\lambda)$ and $\psi(\lambda)$ are relatively prime, it follows that all coefficients c_i and d_i must be zero, and thus R is non-singular.

2. Another lemma. Let $p(\lambda)$ and $\phi(\lambda)$ be given polynomials, of which the first is irreducible and of degree $s \geq 1$. Since there exist at most s polynomials which are linearly independent modulo $p(\lambda)$, it follows that the polynomials $1, \phi, \phi^2, \dots, \phi^s$ are linearly dependent modulo $p(\lambda)$. By a

¹⁰ Turnbull and Aitken, *op. cit.*, chap. 6.

familiar argument, there then exists a unique polynomial $f(x)$, with leading coefficient unity, and of minimum degree, such that

$$f(\phi(\lambda)) \equiv 0 \pmod{p(\lambda)}.$$

It follows readily that $f(x)$ is irreducible, and also that if $g(x)$ is a polynomial, such that $g(\phi(\lambda)) \equiv 0 \pmod{p(\lambda)}$, then $g(x) \equiv 0 \pmod{f(x)}$. We shall let t denote the degree of $f(x)$.

Let ρ denote a root of $p(\lambda) = 0$ in a properly extended field, and consider the three fields,

$$K \subseteq K(\phi(\rho)) \subseteq K(\rho).$$

The field $K(\phi(\rho))$ is seen to be of degree t over K , as $\phi(\rho)$ satisfies the irreducible equation $f(x) = 0$. Also $K(\rho)$ is of degree s over K . Hence, by a well known theorem,¹¹ $K(\rho)$ is algebraic of degree $m = s/t$ over the field $K(\phi(\rho))$. That is, t is a divisor of s , and ρ satisfies no equation of degree less than m with coefficients in $K(\phi(\rho))$. We may now prove the following lemma:

LEMMA 3. *If $F(x, y)$ is a polynomial in the indeterminates x, y , of degree at most $m - 1$ in y , and if*

$$\begin{array}{ll} F(\phi(\lambda), \lambda) \equiv 0 & \pmod{p(\lambda)}, \\ \text{then } F(x, y) \equiv 0 & \pmod{f(x)}. \end{array}$$

Under the hypotheses of the lemma, we have $F(\phi(\rho), \rho) = 0$. Let $F(x, y) = \sum_{i=0}^{m-1} F_i(x) y^i$. We have then $\sum_{i=0}^{m-1} F_i(\phi(\rho)) \rho^i = 0$. But if some $F_i(\phi(\rho)) \neq 0$, this contradicts the fact that ρ can satisfy no equation of degree less than m with coefficients in $K(\phi(\rho))$. Hence $F_i(\phi(\rho)) = 0$, and thus $F_i(\phi(\lambda)) \equiv 0 \pmod{p(\lambda)}$, ($i = 0, 1, \dots, m - 1$). It follows that each $F_i(x) \equiv 0 \pmod{f(x)}$, and the lemma is established. We remark that if the degree of $F(x, y)$ in x is at most $t - 1$, then $F(x, y)$ vanishes identically.

3. The elementary divisors of $\phi(A)$. We come now to the main problem of the paper. Let A be a given matrix, and $\phi(A)$ a given polynomial in A . Since $\phi(HAH^{-1}) = H\phi(A)H^{-1}$, there is no loss of generality in assuming that A is in canonical form. If $A = A_1 \dot{+} A_2$, then $\phi(A) = \phi(A_1) \dot{+} \phi(A_2)$, and the elementary divisors of $\phi(A)$ are precisely those of $\phi(A_1)$, together with those of $\phi(A_2)$. We shall therefore assume henceforth that A has a single elementary divisor $[p(\lambda)]^r$. It follows that the minimum function of A is $[p(\lambda)]^r$. If the degree of $p(\lambda)$ is s , then the order of A is $n = rs$.

¹¹ See, e. g., van der Waerden, *op. cit.*, vol. 1, p. 98.

Let $f(x)$ denote the unique irreducible polynomial of degree t defined in the preceding section. Then we have

$$f(\phi(\lambda)) \equiv 0 \pmod{p(\lambda)}.$$

It may well happen that $f(\phi(\lambda))$ is divisible by a power of $p(\lambda)$ greater than the first. Suppose that it is divisible by $[p(\lambda)]^q$ but not by $[p(\lambda)]^{q+1}$. We now define an integer i as follows. If $q \geq r$, we set $i = r$, while if $q < r$, we place $i = q$. Hence in either case we have $f(\phi(\lambda)) \equiv 0 \pmod{[p(\lambda)]^i}$. We further define positive integers k, l by the relations,

$$(1) \quad r = (k-1)i + l, \quad k \geq 1, \quad 1 \leq l \leq i.$$

It follows that $[f(\phi(\lambda))]^k \equiv 0 \pmod{[p(\lambda)]^r}$, while

$$[f(\phi(\lambda))]^{k-1} \not\equiv 0 \pmod{[p(\lambda)]^r}.$$

The minimum function of $\phi(A)$ is therefore $[f(\lambda)]^k$, and the elementary divisors of $\phi(A)$ are all powers of $f(\lambda)$. If we denote the integer s/t by m , we may state the following precise result:

THEOREM 1. *The matrix $\phi(A)$ has as elementary divisors; $[f(\lambda)]^k$ taken lm times, and $[f(\lambda)]^{k-1}$ taken $m(i-l)$ times.*

We shall prove this theorem by actually exhibiting a matrix R which transforms $\phi(A)$ to canonical form. Let U denote a vector of grade $n = rs$ with respect to A .¹² The R. C. F. of A relative to U is then $[p(\lambda)]^r$.

Let α and β be integers such that $0 \leq \alpha \leq i-1$, $0 \leq \beta \leq m-1$. We shall now make use of Lemma 1, the notation being as in the statement of the lemma, with the exception that we shall find it convenient to replace each subscript j by the two subscripts α and β . That is, e_j becomes $e_{\alpha\beta}$, $h_j(\lambda)$ becomes $h_{\alpha\beta}(\lambda)$, and so on. Two cases will be considered separately.

Case 1. $0 \leq \alpha \leq l-1$, $0 \leq \beta \leq m-1$. Let $e_{\alpha\beta} = tk$, $V_{\alpha\beta} = U[p(A)]^{\alpha A^\beta}$, $B = \phi(A)$. Since $U[p(A)]^{\alpha A^\beta} [f(\phi(A))]^k = 0$, it follows that $h_{\alpha\beta}(\lambda) = [f(\lambda)]^k$, and $Q_{\alpha\beta}$ is therefore the companion matrix of $[f(\lambda)]^k$. The matrix $R_{\alpha\beta}$ has as rows the vectors,

$$(2) \quad U[p(A)]^{\alpha A^\beta}, \quad U[p(A)]^{\alpha A^\beta} \phi(A), \dots, U[p(A)]^{\alpha A^\beta} [\phi(A)]^{tk-1}.$$

Case 2. $l \leq \alpha \leq i-1$, $0 \leq \beta \leq m-1$. In this case, let $e_{\alpha\beta} = t(k-1)$, $V_{\alpha\beta} = U[p(A)]^{\alpha A^\beta}$, $B = \phi(A)$. Since now $f(\phi(\lambda))$ is divisible by $[p(\lambda)]^i$,

¹² If A is in canonical form, we may choose $U = (1, 0, \dots, 0)$, as in § 1.

it follows by relations (1), that $U[p(A)]^{\alpha A^{\beta}}[f(\phi(A))]^{k-1} = 0$, and thus that $h_{\alpha\beta}(\lambda) = [f(\lambda)]^{k-1}$. The matrix $R_{\alpha\beta}$ has as rows the vectors,

$$(3) \quad U[p(A)]^{\alpha A^{\beta}}, \quad U[p(A)]^{\alpha A^{\beta}}\phi(A), \dots, U[p(A)]^{\alpha A^{\beta}}[\phi(A)]^{t(k-1)-1}.$$

It is easily seen that $\sum_{\alpha=0}^{i-1} \sum_{\beta=0}^{m-1} e_{\alpha\beta} = n$, and the hypotheses of Lemma 2 are satisfied. The matrix R then is formed by arranging the matrices $R_{\alpha\beta}$ ($\alpha = 0, 1, \dots, i-1$; $\beta = 0, 1, \dots, m-1$) in some fixed order, and using the same order for the $\xi_{\alpha\beta}$ and $\eta_{\alpha\beta}$ to define ξ and η as in the statement of the lemma. Let us now assume for the present that R is non-singular. We then have $Q = R\phi(A)R^{-1}$, and by the determinations of $h_{\alpha\beta}(\lambda)$ above, we see that Q is the direct sum of the companion matrix of $[f(\lambda)]^k$ taken ml times, and of the companion matrix of $[f(\lambda)]^{k-1}$ taken $m(i-l)$ times. But since $f(\lambda)$ is irreducible, Q is therefore the canonical form of $\phi(A)$, and the elementary divisors are those stated in the theorem.

There remains only to prove that R is non-singular. Any linear combination $UF(\phi(A), A)$ of the row vectors of R (of the types (2) and (3)) corresponds to a polynomial $F(x, y)$ of the form

$$(4) \quad F(x, y) = \sum_{j=0}^{i-1} F_j(x, y) [p(y)]^j,$$

where the degree of $F_j(x, y)$ is at most $m-1$ in y , while its degree in x is at most $tk-1$ for $j=0, 1, \dots, l-1$, and at most $t(k-1)-1$, for $j=l, l+1, \dots, i-1$. If the linear combination of the rows of R is the zero vector, we have

$$UF(\phi(A), A) = 0,$$

and since the R. C. F. of A relative to U is $[p(\lambda)]^r$, it follows that

$$(5) \quad F(\phi(\lambda), \lambda) \equiv 0 \pmod{[p(\lambda)]^r}.$$

We shall complete the proof by showing that under these conditions, all $F_j(x, y)$ vanish identically, and thus the rows of R are linearly independent.

We first dispose of the special case in which $i=r$, and hence $k=1$, $l=r$. In this case, all $F_j(x, y)$ are of degree at most $t-1$ in x . From relation (5), we find that

$$\sum_{j=0}^{r-1} F_j(\phi(\lambda), \lambda) [p(\lambda)]^j \equiv 0 \pmod{[p(\lambda)]^r}.$$

Now clearly $F_0(\phi(\lambda), \lambda) \equiv 0 \pmod{p(\lambda)}$, and by Lemma 3, it follows that $F_0(x, y) \equiv 0 \pmod{f(x)}$. But being of degree at most $t-1$ in x , $F_0(x, y)$

must vanish identically. We now pass on to $F_1(x, y)$, and a similar argument shows that it is also identically zero. A continuation of this process establishes the fact that all $F_j(x, y)$ vanish identically.

Suppose now that $i < r$. By definition of i , we know that $f(\phi(\lambda))$ is then divisible by $[p(\lambda)]^i$ but not by $[p(\lambda)]^{i+1}$. We now assume that all $F_j(x, y)$ are divisible by $[f(x)]^\gamma$ where $0 \leq \gamma < k-1$, and shall show that they are all divisible by $[f(x)]^{\gamma+1}$. If we set $F_j(x, y) = [f(x)]^\gamma F'_j(x, y)$, we get from (5),

$$(6) \quad \sum_{j=0}^{i-1} F'_j(\phi(\lambda), \lambda) [p(\lambda)]^j \equiv 0 \pmod{[p(\lambda)]^{r-\gamma i}}.$$

Clearly $F'_0(\phi(\lambda), \lambda) \equiv 0 \pmod{p(\lambda)}$, and by Lemma 3 we have $F'_0(x, y) \equiv 0 \pmod{f(x)}$. Suppose that $F'_j(x, y) \equiv 0 \pmod{f(x)}$, ($j = 0, 1, \dots, \delta$) where $0 \leq \delta < i-1$. Since $r - \gamma i > i$, it follows that $F'_{\delta+1}(\phi(\lambda), \lambda) \equiv 0 \pmod{p(\lambda)}$, and thus $F'_{\delta+1}(x, y) \equiv 0 \pmod{f(x)}$. Hence $F'_j(x, y) \equiv 0 \pmod{f(x)}$, ($j = 0, 1, \dots, i-1$). It therefore follows that all $F_j(x, y)$ are divisible by $[f(x)]^{\gamma+1}$, and a process of induction then shows that they are all divisible by $[f(x)]^{k-1}$. But the $F_j(x, y)$ ($j = l, l+1, \dots, i-1$) are of degree at most $t(k-1) - 1$ in x , and hence must vanish identically.

Now let $F_j(x, y) = [f(x)]^{k-1} F^*_j(x, y)$, ($j = 0, 1, \dots, l-1$). From relation (5) we then have

$$(7) \quad \sum_{j=0}^{l-1} F^*_j(\phi(\lambda), \lambda) [p(\lambda)]^j \equiv 0 \pmod{[p(\lambda)]^i}.$$

A repetition of the argument of the preceding paragraphs shows that each $F^*_j(x, y)$ is divisible by $f(x)$, and thus $F_j(x, y)$ is divisible by $[f(x)]^k$, ($j = 0, 1, \dots, l-1$). But these $F_j(x, y)$ are of degree at most $tk - 1$ in x , and must therefore vanish identically. This completes the proof of the theorem.

Examples. Let K be an algebraically closed field, and suppose A has the single elementary divisor $(\lambda - a)^n$. Then in terms of our notation, we have $p(\lambda) = \lambda - a$, $s = 1$, $r = n$. Let now $\phi(\lambda)$ be expanded in powers of $\lambda - a$,

$$\phi(\lambda) = a_0 + a_1(\lambda - a) + a_2(\lambda - a)^2 + \dots$$

Then clearly $\phi(\lambda) - a_0 \equiv 0 \pmod{(\lambda - a)}$, so that $f(x) = x - a_0$, and $t = 1$, $m = s = 1$. If now the first number of the sequence $a_1, a_2, \dots, a_{n-1}, 1$, which is not zero is the i -th, then we have

$$f(\phi(\lambda)) \equiv 0 \pmod{(\lambda - a)^i},$$

while if $i < n$,

$$f(\phi(\lambda)) \not\equiv 0 \pmod{(\lambda - a)^{i+1}}.$$

Thus this definition of i corresponds to that given in the notation above, and if we define k and l by the relations (1), our theorem tells us that $\phi(A)$ has the elementary divisor $(\lambda - a_0)^k$ taken l times and $(\lambda - a_0)^{k-1}$ taken $(i - l)$ times. Thus our general theorem reduces to the one obtained previously for this case by the writers referred to in the introduction.

As a second example, let K be the field of real numbers, and A a matrix of order $n = 6$, with the elementary divisor $(\lambda^2 + 1)^3$. We may take A in the canonical form,

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & -3 & 0 & -3 & 0 \end{pmatrix}.$$

Then $p(\lambda) = \lambda^2 + 1$, $s = 2$, $r = 3$. Suppose $\phi(\lambda) = \lambda^3 + 3\lambda$. Then $\phi(\lambda) \equiv 2\lambda \pmod{p(\lambda)}$, $[\phi(\lambda)]^2 \equiv -4 \pmod{p(\lambda)}$, and hence $f(x) = x^2 + 4$. We have then $t = 2$, $m = 1$. It is easily verified that $\phi^2 + 4 \equiv 0 \pmod{[p(\lambda)]^2}$, but $\not\equiv 0 \pmod{[p(\lambda)]^3}$. Hence $i = 2$, $k = 2$, $l = 1$. Our theorem then states that $\phi(A) = A^3 + 3A$ has the elementary divisors $(\lambda^2 + 4)^2$, $\lambda^2 + 4$. Thus the canonical form of $\phi(A)$ is the matrix

$$Q = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ -16 & 0 & -8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -4 & 0 \end{pmatrix}.$$

The vector $U = (1, 0, 0, 0, 0, 0)$ is of grade 6 relative to A , and so the matrix R is a matrix whose rows are respectively U , $U\phi(A)$, $U[\phi(A)]^2$, $U[\phi(A)]^3$, $U(A^2 + 1)$, $U(A^2 + 1)\phi(A)$. A calculation shows that

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 1 & 0 & 0 \\ -1 & 0 & 6 & 0 & 3 & 0 \\ 0 & -6 & 0 & 8 & 0 & 6 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 4 & 0 & 1 \end{pmatrix}.$$

It is easily verified that $R\phi(A) = QR$, and hence that $Q = R\phi(A)R^{-1}$.

4. Solution of matrix equations. Let B be a given matrix of order n , and $\phi(\lambda)$ a given polynomial in the scalar variable λ . It will be understood that all elements and operations are to be restricted to the given field K . We shall now give a brief account of an application of the results of the preceding section to the solution of the equation,

$$(8) \quad \phi(X) = B,$$

where X is a matrix of order n to be determined. A different method of solving this equation has recently been given by Ingraham.¹³

If S is a non-singular matrix, then $\phi(SXS^{-1}) = S\phi(X)S^{-1} = SBS^{-1}$. Hence there is no loss of generality in assuming that B is in canonical form. We shall assume henceforth that B is in canonical form, and is therefore the direct sum of the companion matrices of its elementary divisors. We observe that, if X is a solution of the equation (8), then SXS^{-1} is also a solution, if and only if S is commutative with B .

We shall consider first the case in which the elementary divisors of B are all powers of a single irreducible polynomial $f(\lambda)$. Let

$$(9) \quad f(\phi(\lambda)) = a[p_1(\lambda)]^{n_1}[p_2(\lambda)]^{n_2} \cdots [p_k(\lambda)]^{n_k},$$

be the decomposition of $f(\phi(\lambda))$ into powers of its distinct irreducible factors, each with leading coefficient unity. It follows easily that, if $p(\lambda)$ is any irreducible polynomial, with leading coefficient unity, then $f(x)$ is the unique minimum polynomial (defined in § 2) such that

$$f(\phi(\lambda)) \equiv 0 \pmod{p(\lambda)},$$

if and only if $p(\lambda)$ is one of the $p_i(\lambda)$ occurring in (9).

Let X denote a solution of the equation (8), and Y the canonical form of X . That is, $Y = Y_1 \dot{+} Y_2 \dot{+} \cdots \dot{+} Y_r$, where the Y_i are the companion matrices of the elementary divisors of X . Then

$$\phi(Y) = \phi(Y_1) \dot{+} \phi(Y_2) \dot{+} \cdots \dot{+} \phi(Y_r)$$

is similar to B , and the elementary divisors of B are precisely the elementary divisors of all the $\phi(Y_i)$. But by the results of the preceding section, $\phi(Y_i)$ can have elementary divisors which are powers of $f(\lambda)$, if and only if Y_i is the companion matrix of some power of a $p_i(\lambda)$ occurring in (9). Suppose then that the elementary divisors of Y are

¹³ M. H. Ingraham, "On the rational solutions of the matrix equation $P(X) = A$," *Journal of Mathematics and Physics*, vol. 13 (1934), pp. 46-50. For additional references to matrix equations see MacDuffee, *op. cit.*, chap. 8.

$$(10) \quad \begin{array}{ccccccc} [p_1(\lambda)]^{n_{11}}, & \cdots, & [p_1(\lambda)]^{n_{1t_1}}, \\ [p_2(\lambda)]^{n_{21}}, & \cdots, & [p_2(\lambda)]^{n_{2t_2}}, \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ [p_k(\lambda)]^{n_{k1}}, & \cdots, & [p_k(\lambda)]^{n_{kt_k}}, \end{array}$$

where $n_{ij} \geq n_{il}$ if $l > j$. Let the degree of $p_i(\lambda)$ be denoted by N_i . Then we must have

$$(11) \quad (n_{11} + n_{12} + \cdots + n_{1t_1})N_1 + \cdots + (n_{k1} + n_{k2} + \cdots + n_{kt_k})N_k = n.$$

We are now in a position to give a method of finding all solutions of the equation (8). Form the diophantine equation (11), and solve it for the n_{ij} under the condition that $n_{ij} \geq n_{il}$ if $l > j$. Each solution gives us the elementary divisors (10) of a matrix, which is a possible solution of our equation. Form the matrix Y , which is the direct sum of the companion matrices of these elementary divisors. Then by Theorem 1, it is easy to find the elementary divisors of $\phi(Y)$. If these elementary divisors are not the same as the elementary divisors of B , this Y is discarded. If, however, the elementary divisors are identical, then $\phi(Y)$ is similar to B , and the proof of Theorem 1 shows how to find a matrix R such that $R\phi(Y)R^{-1} = B$. If we let $X = RYR^{-1}$, then X is a solution of the equation (8). If X_1, X_2, \cdots, X_q is a complete set of dissimilar solutions, all of which can be found by this method, then the most general solutions are of the form LX_iL^{-1} , where L is a non-singular matrix commutative with B .

It is not difficult to write out additional equations, which together with the equation (11) will serve to determine completely the admissible matrices Y , but the tentative procedure outlined above is perhaps as easy to apply in any given case.

We now return to the general case in which the elementary divisors of B are unrestricted. Suppose these elementary divisors are

$$(12) \quad \begin{array}{ccccccc} [f_1(\lambda)]^{m_{11}}, & \cdots, & [f_1(\lambda)]^{m_{1r_1}}, \\ [f_2(\lambda)]^{m_{21}}, & \cdots, & [f_2(\lambda)]^{m_{2r_2}}, \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ [f_l(\lambda)]^{m_{l1}}, & \cdots, & [f_l(\lambda)]^{m_{lr_l}}, \end{array}$$

where the $f_i(\lambda)$ ($i = 1, 2, \cdots, l$) are distinct and irreducible. We may then write $B = B_1 + B_2 + \cdots + B_l$, where B_i is the direct sum of the companion matrices of the elementary divisors occurring in the i -th row of the table (12). We shall now prove the following theorem:

THEOREM 2. If X is any solution of the equation (8), then

$$X = X_1 \dot{+} X_2 + \cdots \dot{+} X_l,$$

where X_i is of the same order as B_i , and is a solution of the equation, $\phi(X) = B_i$ ($i = 1, 2, \cdots, l$).

Let

$$(13) \quad f_i(\phi(\lambda)) = b_i[p_{i1}(\lambda)]^{a_{i1}}[p_{i2}(\lambda)]^{a_{i2}} \cdots [p_{is_i}(\lambda)]^{a_{is_i}}, \\ (i = 1, 2, \cdots, l)$$

be the decomposition of the $f_i(\phi(\lambda))$ into powers of distinct irreducible factors, each with leading coefficient unity. If X is a given solution of equation (8), it follows by an argument similar to that used above that the elementary divisors of X are all powers of the $p_{ij}(\lambda)$ ($i = 1, 2, \cdots, l$; $j = 1, 2, \cdots, s_i$). Let Y_i denote the direct sum of the companion matrices of the elementary divisors of X which are powers of the functions $p_{i1}(\lambda), \cdots, p_{is_i}(\lambda)$ ($i = 1, 2, \cdots, l$), and set $Y = Y_1 \dot{+} Y_2 \dot{+} \cdots \dot{+} Y_l$. Since the $f_i(\lambda)$ are distinct, it follows that the $p_{ij}(\lambda)$ are all distinct, and the elementary divisors of $\phi(Y)$ which are powers of $f_i(\lambda)$ are precisely the elementary divisors of $\phi(Y_i)$. Hence Y_i is of the same order as B_i , and $\phi(Y_i)$ is similar to B_i ($i = 1, 2, \cdots, l$). Let us set $SXS^{-1} = Y$, $T_i\phi(Y_i)T_i^{-1} = B_i$, $T = T_1 \dot{+} T_2 \dot{+} \cdots \dot{+} T_l$. We have then $T\phi(Y)T^{-1} = B$, from which it follows that $\phi(TSXS^{-1}T^{-1}) = B$, and thus TS is commutative with B . It is then known¹⁴ that TS is of the form $M_1 \dot{+} M_2 \dot{+} \cdots \dot{+} M_l$, where M_i is of the same order as B_i , and is commutative with B_i . A calculation shows that

$$X = M_1^{-1}T_1Y_1T_1^{-1}M_1 \dot{+} \cdots \dot{+} M_l^{-1}T_lY_lT_l^{-1}M_l.$$

We find also that $\phi(M_i^{-1}T_iY_iT_i^{-1}M_i) = M_i^{-1}T_i\phi(Y_i)T_i^{-1}M_i = M_i^{-1}B_iM_i = B_i$. The theorem is therefore established.

By means of this theorem, the solution of the general equation (8) is seen to reduce to the solution of a set of equations of the comparatively simple type, in which the elementary divisors of B are all powers of a single irreducible polynomial. Thus all solutions can be found by the method discussed earlier in this section.

SMITH COLLEGE,
NORTHAMPTON, MASS.

¹⁴ See O. Schreier and B. L. van der Waerden, "Die Automorphismen der projektiven Gruppen," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, vol. 6 (1928), p. 308.

ON CERTAIN TYPES OF HEXAGONS.¹

By J. R. MUSSELMAN.

1. The resolvent, $V_1 = x_0 + \epsilon x_1 + \epsilon^2 x_2 + \cdots + \epsilon^{n-1} x_{n-1}$, where ϵ is a primitive n -th root of unity, was introduced by Lagrange² in his memoirs devoted to the fundamental principles of the solutions of the cubic and quartic equations. Its entrance, however, into the field of geometry is very recent. If we represent any point P in the plane by the single complex number p , and if M_k ($k = 0, 1, \cdots, n-1$) represents the n vertices of a positively-ordered polygon, then when the coördinates p_k of these vertices are subject to one and only one condition, namely that

$$V_1 = \sum_{k=0}^{n-1} \epsilon^k p_k = 0$$

we obtain a polygon which we shall call a positive n -gon of type M . R. L. Echols³ has used these polygons in giving geometric pictures of the solutions of the cubic and quartic equations. The writer⁴ has pointed out recently two different constructions in which these n -gons of type M occur.

In addition to the above studies of this particular type of n -gon, the Lagrange resolvents (and their conjugates) have been used by L. M. Blumenthal⁵ to prove that the norm-area of a $2n$ -gon is unaltered by translating either of its component n -gons. The Morleys⁶ in their recent book have shown that under homologies the $n-1$ Lagrange resolvents for a n -gon form a complete system of relative invariants, and have used them in considering some special ordered n -points. In this article, the Lagrange resolvents are used to disclose some new facts about a well-known figure, to characterize certain interesting ordered six-points, and to prove that connected with any

¹ Read before the National Academy of Science, November 19, 1934.

² *Memoirs of Berlin Academy*, 1769; reprinted in *Oeuvres de Lagrange* (Paris, 1868), vol. 3, p. 207.

³ *The Roots of Circulants and Application to the Roots of Polynomials*. The University of Virginia (1928).

⁴ "On certain types of polygons," *The American Mathematical Monthly*, vol. 40 (1933), p. 157.

⁵ "Lagrange resolvents in Euclidean geometry," *The American Journal of Mathematics*, vol. 49 (1927), p. 511.

⁶ *Inversive Geometry*, G. Bell & Sons, London (1933), p. 203.

six points there are two circumscribed hexagons, whose opposite sides are parallel and whose vertices lie on rectangular hyperbolas.

2. If on the sides of any triangle $A_1A_2A_3$ we construct the positively ordered equilateral triangles $A_1A_3A_{13}$, $A_2A_1A_{21}$, and $A_3A_2A_{32}$ the coördinates of the three vertices are $A_{13}(-\omega a_1 - \omega^2 a_3)$, $A_{21}(-\omega a_2 - \omega^2 a_1)$, and $A_{32}(-\omega a_3 - \omega^2 a_2)$ where $\omega^3 = 1$. The vector $A_{32}A_1$ is the Lagrange resolvent $a_1 + \omega^2 a_2 + \omega a_3$, which we shall term u_2 . Similarly, the vectors $A_{13}A_2$ and $A_{21}A_3$ are ωu_2 and $\omega^2 u_2$. Hence, we have the well-known theorem that the Lagrange resolvents $A_{32}A_1$, $A_{13}A_2$ and $A_{21}A_3$ are equal in length and intersect at angles of $2\pi/3$. These vectors meet at a point f_2 whose coördinate is

$$(2.1) \quad f_2 = g - u_2 \bar{u}_1 / 3 \bar{u}_2. \quad g = (a_1 + a_2 + a_3) / 3$$

Similarly, if we construct on the sides of the triangle $A_1A_2A_3$ the positively-ordered equilateral triangles $A_1A_2A_{12}$, $A_2A_3A_{23}$, and $A_3A_1A_{31}$ then the vectors $A_{23}A_1$, $A_{31}A_2$ and $A_{12}A_3$ are respectively u_1 , ωu_1 and $\omega^2 u_1$ where u_1 is the Lagrange resolvent $a_1 + \omega a_2 + \omega^2 a_3$. These vectors meet at the point f_1 whose coördinate is

$$(2.2) \quad f_1 = g - u_1 \bar{u}_2 / 3 \bar{u}_1.$$

The points f_1 and f_2 are variously known as the Fermat points⁷ of the triangle $A_1A_2A_3$ or as the isogenic centers.⁸

The area of the triangle $A_{13}A_{21}A_{32}$ is five-halves that of $A_1A_2A_3$ plus $3\frac{1}{2}s^2/8$, while the area of $A_{31}A_{12}A_{23}$ is five-halves that of $A_1A_2A_3$ minus $3\frac{1}{2}s^2/8$ where $s^2 = \overline{A_1A_2}^2 + \overline{A_2A_3}^2 + \overline{A_3A_1}^2$. Hence, the sum of the areas of the triangles $A_{13}A_{21}A_{32}$ and $A_{31}A_{12}A_{23}$ is 5 times the area of the triangle $A_1A_2A_3$. The hexagons $A_{12}A_{21}A_{32}A_{13}A_{31}A_{23}$, $A_{12}A_{31}A_{13}A_{21}A_{32}A_{23}$ and $A_{12}A_{31}A_{23}A_{32}A_{13}A_{21}$ are n -gons of type M , i. e. hexagons for which the Lagrange resolvent V_1 vanishes. Their areas are respectively $(\overline{A_1A_2}^2 + \overline{A_3A_1}^2)3\frac{1}{2}/4$, $(\overline{A_2A_3}^2 + \overline{A_3A_1}^2)3\frac{1}{2}/4$ and $(\overline{A_1A_2}^2 + \overline{A_2A_3}^2)3\frac{1}{2}/4$. In terms of their five Lagrange resolvents these hexagons can be characterized respectively as $V_1 = 2V_2 + \omega V_4 = V_3 - 2\omega V_5 = 0$; $V_1 = 2V_2 + V_4 = V_3 - 2V_5 = 0$; $V_1 = 2V_2 + \omega^2 V_4 = V_3 - 2\omega^2 V_5 = 0$.

The hexagon $A_{23}A_{21}A_{31}A_{32}A_{12}A_{13}$ is worth some attention. Its area is twice that of the triangle $A_1A_2A_3$ and in terms of its resolvents we find $V_3 = V_1 - 3V_4 = 3V_2 + V_5 = 0$. To discover the geometrical significance of these conditions it is essential to express the resolvents of the hexagon in

⁷ Morley, *loc. cit.*, p. 207.

⁸ R. A. Johnson, *Modern Geometry*, p. 218.

terms of the resolvents of its two component triangles. Thus, if we denote by u_1 and u_2 the two Lagrange resolvents of the triangle $A_{23}A_{31}A_{12}$, and by u'_1 and u'_2 those for the triangle $A_{32}A_{13}A_{21}$ we can easily prove that *the necessary and sufficient conditions for a positively-ordered hexagon to have $V_3 = V_1 - 3V_4 = 3V_2 + V_5 = 0$ are that the centroids of its component triangles coincide, that the vector u'_1 be negatively parallel to u_1 and half its length, and that the vector u'_2 be negatively parallel to u_2 and twice its length.*

If we denote by $f_1, f_2; f'_1, f'_2; F_1, F_2$ the Fermat points of the triangles $A_{23}A_{31}A_{12}, A_{32}A_{13}A_{21}, A_1A_2A_3$ respectively; and by $h_1, h_2; h'_1, h'_2; H_1, H_2$ the Hessian points of the same triangles, then the following facts can be verified—the three triangles have the same centroid; F_1 coincides with f_2 and F_2 with f'_1 ; g, f_1, f'_1, h_2, h'_2 and H_1 lie on a line and so do g, f_2, f'_2, h_1, h'_1 and H_2 . The distances between these points can be readily read from the relations

$$(2.3) \quad \begin{aligned} g - f'_2 &= 2(f_2 - g); & g - f_1 &= 2(f'_1 - g) \\ g - H_1 &= 2(h_2 - g); & g - H_2 &= 2(h'_1 - g) \\ g - h_1 &= 4(g - H_2); & g - h'_2 &= 4(g - H_1). \end{aligned}$$

3. In this section, let us consider, in terms of their Lagrange resolvents, some special ordered six-points which possess features of interest. We shall first prove the theorem that

The necessary and sufficient condition for a positively-ordered hexagon to have $V_1 = V_2 = 0$ is that the sides of the triangle $x_4x_6x_2$ form positive right angles with the corresponding medians of the triangle $x_1x_3x_5$ and equal $2.3^{-1/2}$ times their length.

Since $V_1 = V_2 = 0$, we have

$$\begin{aligned} x_1 + \omega x_3 + \omega^2 x_5 &= x_4 + \omega x_6 + \omega^2 x_2 \\ x_1 + \omega^2 x_3 + \omega x_5 &= -x_4 - \omega^2 x_6 - \omega x_2 \end{aligned}$$

whence by addition,

$$(3.1) \quad \begin{aligned} 3x_1 - 3g &= 3^{1/2}i(x_6 - x_2), & 3g &= x_1 + x_3 + x_5 \\ \text{or } x_2 - x_6 &= 3^{1/2}i(x_1 - g). \end{aligned}$$

Similarly, we can show that

$$(3.2) \quad \begin{aligned} x_4 - x_2 &= 3^{1/2}i(x_3 - g) \\ \text{and } x_6 - x_4 &= 3^{1/2}i(x_5 - g) \end{aligned}$$

which demonstrates the theorem. The conditions can easily be shown to be sufficient. Now if g' be the centroid of the triangle $x_4x_6x_2$ one can show that

$$(3.3) \quad \begin{aligned} x_3 - x_1 &= 3^{1/2}i(x_2 - g') \\ x_5 - x_3 &= 3^{1/2}i(x_4 - g') \\ x_1 - x_5 &= 3^{1/2}i(x_6 - g') \end{aligned}$$

so that the sides of the triangle $x_1x_3x_5$ are perpendicular to the corresponding medians of the triangle $x_4x_6x_2$ and equal to $2.3^{-1/2}$ times their length. Thus, we have a mutual relationship between the two triangles.⁹ In addition, since

$$\begin{vmatrix} x_4 & x_6 & x_2 \\ \bar{x}_4 & \bar{x}_6 & \bar{x}_2 \\ 1 & 1 & 1 \end{vmatrix} = 3 \begin{vmatrix} x_1 & x_3 & g \\ \bar{x}_1 & \bar{x}_3 & \bar{g} \\ 1 & 1 & 1 \end{vmatrix}$$

we see that the area of the triangle $x_4x_6x_2$ is equivalent to that of $x_1x_3x_5$. Also

$$\begin{aligned} f_2 - g &= f'_2 - g', & f_1 - g &= g' - f'_1 \\ h_1 - g &= h'_1 - g', & h_2 - g &= g' - h'_2 \end{aligned}$$

whence the vector $f_2 - g$ is equal and positively parallel to $f'_2 - g'$, but $f_1 - g$ is equal and negatively parallel to $f'_1 - g'$; etc.

From the formulae (3.1), (3.2), and (3.3), we read that if *perpendiculars, dropped from the vertices x_4, x_6, x_2 of a triangle to the corresponding sides of the triangle $x_1x_3x_5$, should meet at the centroid of $x_4x_6x_2$, then the perpendiculars from the vertices x_1, x_3, x_5 to the sides of $x_4x_6x_2$ will meet at the centroid of $x_1x_3x_5$.*

The necessary and sufficient condition for a positively-ordered six-point to have $V_4 = V_5 = 0$ is that the sides of the triangle $x_4x_6x_2$ form negative right angles with the corresponding medians of the triangle $x_1x_3x_5$ and equal $2.3^{-1/2}$ times their length. This relationship is mutual and again both triangles are equivalent in area. If in addition $V_3 = 0$, both centroids coincide, and both Fermat points f_1 and f'_1 , hence the diagonals of the hexagon meet at angles of $2\pi/3$.

The necessary and sufficient condition for a positively-ordered six-point to have $V_2 = V_4 = 0$ is that the midpoint of each of its diagonals should be the midpoint of the centroids of the triangles $x_1x_3x_5$ and $x_4x_6x_2$. The two triangles have their corresponding sides equal and parallel; they are inversely equivalent in area and also perspective. The opposite sides of the hexagon are equal and negatively parallel. Also

$$\begin{aligned} f'_2 - g' &= g - f_2; & f'_1 - g' &= g - f_1 \\ h_1 - g &= g - h'_1; & h_2 - g &= g - h'_2. \end{aligned}$$

⁹ If $V_1 = V_2 = V_3 = 0$, we have the special case of the above, in which the centroids g and g' coincide. See Morley, *loc. cit.*, p. 214 for details.

Hence, the join of f'_2 and f'_1 is parallel to f_1 and f_2 ; also the join of h'_2 and h'_1 to that of h_1 and h_2 . If in addition $V_3 = 0$, we can construct the component triangles as follows—starting with the triangle $x_4x_6x_2$ with centroid g' , then x_1 lies on the median from x_4 such that $x_4x_1 = 2x_4g'$; similarly for x_3 and x_5 .

The necessary and sufficient condition for a positively ordered six-point to have $V_1 = V_5 = 0$ is that each diagonal shall be parallel and equal in length to the vector joining the centroids g and g' of the two component triangles. These two triangles have their corresponding sides equal and parallel and are directly congruent. If in addition, we make $V_3 = 0$, then the two triangles will coincide throughout, and the hexagon is a doubly-counted triangle.

4. Let A_k ($k = 1, 2, \dots, 6$) be any positively-ordered hexagon and let us construct the following six positive hexagons for which the Lagrange resolvent V_1 vanishes, $P'_1A_2A_3A_4A_5A_6$, $P'_2A_3A_4A_5A_6A_1$, $P'_3A_4A_5A_6A_1A_2$, $P'_4A_5A_6A_1A_2A_3$, $P'_5A_6A_1A_2A_3A_4$ and $P'_6A_1A_2A_3A_4A_5$. The coordinates of the points P'_i ($i = 1, 2, \dots, 6$) are respectively $a_1 - V_1$, $\omega V_1 + a_2$, $a_3 - \omega^2 V_1$, $V_1 + a_4$, $a_5 - \omega V_1$ and $\omega^2 V_1 + a_6$. The equations of the six lines $A_iP'_i$ are

$$\begin{aligned}\bar{V}_1x - V_1\bar{x} + V_1\bar{a}_1 - a_1\bar{V}_1 &= 0 \\ \bar{V}_1x - \omega^2 V_1\bar{x} + \omega^2 V_1\bar{a}_2 - a_2\bar{V}_1 &= 0 \\ \bar{V}_1x - \omega V_1\bar{x} + \omega V_1\bar{a}_3 - a_3\bar{V}_1 &= 0 \\ \bar{V}_1x - V_1\bar{x} + V_1\bar{a}_4 - a_4\bar{V}_1 &= 0 \\ \bar{V}_1x - \omega^2 V_1\bar{x} + \omega^2 V_1\bar{a}_5 - a_5\bar{V}_1 &= 0 \\ \bar{V}_1x - \omega V_1\bar{x} + \omega V_1\bar{a}_6 - a_6\bar{V}_1 &= 0.\end{aligned}$$

From the form of these equations, they represent three pairs of parallel lines, also each line makes a positively-directed angle of $2\pi/3$ with the consecutive line. The coordinates of the point of intersection of each line with the consecutive line are

$$\begin{aligned}P_1: & V_1(\bar{a}_6 - \bar{a}_1) + \bar{V}_1(a_1 - \omega^2 a_6) \div (1 - \omega^2) \bar{V}_1 \\ P_2: & \omega^2 V_1(\bar{a}_1 - \bar{a}_2) + \bar{V}_1(a_2 - \omega^2 a_1) \div \quad \quad \quad \text{“} \\ P_3: & \omega V_1(\bar{a}_2 - \bar{a}_3) + \bar{V}_1(a_3 - \omega^2 a_2) \div \quad \quad \quad \text{“} \\ P_4: & V_1(\bar{a}_3 - \bar{a}_4) + \bar{V}_1(a_4 - \omega^2 a_3) \div \quad \quad \quad \text{“} \\ P_5: & \omega^2 V_1(\bar{a}_4 - \bar{a}_5) + \bar{V}_1(a_5 - \omega^2 a_4) \div \quad \quad \quad \text{“} \\ P_6: & \omega V_1(\bar{a}_5 - \bar{a}_6) + \bar{V}_1(a_6 - \omega^2 a_5) \div \quad \quad \quad \text{“}.\end{aligned}$$

If we call the join of the lines $\overline{A_1P'_1}$ and $\overline{A_3P'_3}$ by B_5 ; of $\overline{A_3P'_3}$ and $\overline{A_5P'_5}$ by B_1 ; of $\overline{A_5P'_5}$ and $\overline{A_1P'_1}$ by B_3 ; of $\overline{A_6P'_6}$ and $\overline{A_2P'_2}$ by B_4 ; of $\overline{A_2P'_2}$ and $\overline{A_4P'_4}$ by B_6 ; of $\overline{A_4P'_4}$ and $\overline{A_6P'_6}$ by B_2 then one can show that $B_1B_5B_3$ and

$B_4B_2B_6$ are equal positive equilateral triangles with corresponding sides positively parallel. Now the necessary and sufficient condition that the six points of intersection of two parallel equilateral triangles—sides produced if necessary—lie on a rectangular hyperbola¹⁰ is that the sides of the two triangles be equal. Consequently, the six points of intersection of the triangles $B_1B_5B_3$ and $B_4B_2B_6$, which are the six points P_i ($i = 1, 2, \dots, 6$), lie on a rectangular hyperbola. In terms of the Lagrange resolvents this six-point is characterized by $V_1 = V_3 = V_5\bar{V}_5 - 4V_3\bar{V}_3 = 0$. Hence, the sides of the triangle $P_4P_6P_2$ make positive right angles with the corresponding medians of $P_1P_3P_5$ and are equal to $2.3^{-\frac{1}{2}}$ times their length; also the Lagrange resolvent u_2 of the triangle $P_1P_3P_5$ is three times the length of the join of the two centroids.

Again, if A_k be any positively-ordered hexagon and we construct the six positive hexagons $P'_1A_2A_3A_4A_5A_6, \dots$ for which the resolvent V_5 vanishes, we will obtain by a process similar to the above-mentioned one a six-point P_i for which $V_4 = V_6 = V_1\bar{V}_1 - 4V_3\bar{V}_3 = 0$. Hence, the sides of the triangle $P_4P_6P_2$ form negative right angles with the corresponding medians of $P_1P_3P_5$ and are equal to $2.3^{-\frac{1}{2}}$ times their length, also the Lagrange resolvent u_1 of the triangle $P_1P_3P_5$ is three times the length of the join of the two centroids. The points $B_1B_5B_3$ and $B_4B_6B_2$ are equal positive equilateral triangles with corresponding sides positively parallel and therefore the vertices of this hexagon lie on a rectangular hyperbola. Hence, associated with any positively-ordered hexagon are two circumscribed six-points, whose opposite sides are parallel and whose vertices lie on rectangular hyperbolas.

However, if we construct the six positive hexagons $P'_1A_2A_3A_4A_5A_6, \dots$ for which the resolvent V_2 vanishes, we will obtain a six-point P_i for which $V_2 = V_4 = V_1\bar{V}_1 - 4V_3\bar{V}_3 = 0$. The points $B_1B_3B_5$ and $B_4B_6B_2$ are equal positive equilateral triangles with corresponding sides negatively parallel. Similarly, if we construct the six positive hexagons for which the resolvent V_4 vanishes, we will obtain a six-point P_i for which $V_2 = V_4 = V_5\bar{V}_5 - 4V_3\bar{V}_3 = 0$. The points $B_1B_5B_3$ and $B_4B_2B_6$ are equal positive equilateral triangles with corresponding sides negatively parallel. Hence, associated with any positively-ordered hexagon are two circumscribed six-points whose opposite sides are equal and parallel, whose diagonals pass through a point, and whose vertices lie on conics.

WESTERN RESERVE UNIVERSITY.

¹⁰ J. R. Musselman, *The American Mathematical Monthly*, vol. 41 (1934), p. 634.

ON THE ABSTRACT PROPERTIES OF LINEAR DEPENDENCE.¹

By HASSLER WHITNEY.

1. **Introduction.** Let C_1, C_2, \dots, C_m be the columns of a matrix M . Any subset of these columns is either linearly independent or linearly dependent; the subsets thus fall into two classes. These classes are not arbitrary; for instance, the two following theorems must hold:

(a) Any subset of an independent set is independent.

(b) If N_p and N_{p+1} are independent sets of p and $p + 1$ columns respectively, then N_p together with some column of N_{p+1} forms an independent set of $p + 1$ columns.

There are other theorems not deducible from these; for in § 16 we give an example of a system satisfying these two theorems but not representing any matrix. Further theorems seem, however, to be quite difficult to find. Let us call a system obeying (a) and (b) a "matroid." The present paper is devoted to a study of the elementary properties of matroids. The fundamental question of completely characterizing systems which represent matrices is left unsolved. In place of the columns of a matrix we may equally well consider points or vectors in a Euclidean space, or polynomials, etc.

This paper has a close connection with a paper by the author on linear graphs;² we say a subgraph of a graph is independent if it contains no circuit. Although graphs are, abstractly, a very small subclass of the class of matroids, (see the appendix), many of the simpler theorems on graphs, especially on non-separable and dual graphs, apply also to matroids. For this reason, we carry over various terms in the theory of graphs to the present theory. Remarkably enough, for matroids representing matrices, dual matroids have a simple geometrical interpretation quite different from that in the case of graphs (see § 13).

The contents of the paper are as follows: In Part I, definitions of matroids in terms of the concepts rank, independence, bases, and circuits are considered, and their equivalence shown. Some common theorems are deduced (for instance Theorem 8). Non-separable and dual matroids are studied in

¹ Presented to the American Mathematical Society, September, 1934.

² "Non-separable and planar graphs," *Transactions of the American Mathematical Society*, vol. 34 (1932), pp. 339-362. We refer to this paper as G.

Part II; this section might replace much of the author's paper G. The subject of Part III is the relation between matroids and matrices. In the appendix, we completely solve the problem of characterizing matrices of integers modulo 2, of interest in topology.

I. MATROIDS.

2. Definitions in terms of rank. Let a set M of elements e_1, e_2, \dots, e_n be given. Corresponding to each subset N of these elements let there be a number $r(N)$, the *rank* of N . If the three following postulates are satisfied, we shall call this system a *matroid*.

(R₁) *The rank of the null subset is zero.*

(R₂) *For any subset N and any element e not in N ,*

$$r(N + e) = r(N) + k, \quad (k = 0 \text{ or } 1).$$

(R₃) *For any subset N and elements e_1, e_2 not in N , if $r(N + e_1) = r(N + e_2) = r(N)$, then $r(N + e_1 + e_2) = r(N)$.*

Evidently *any subset of a matroid is a matroid*. In what follows, M is a fixed matroid. We make the following definitions:

$$\rho(N) = \text{number of elements in } N.$$

$$n(N) = \rho(N) - r(N) = \text{nullity of } N.$$

N is *independent*, or, the elements of N are independent, if $n(N) = 0$; otherwise, N , and its set of elements, are *dependent*.

LEMMA 1. *For any N , $r(N) \geq 0$ and $n(N) \geq 0$. If $N \subset M$, then $r(N) \leq r(M)$, $n(N) \leq n(M)$.*

LEMMA 2. *Any subset of an independent set is independent.*

e is *dependent on* N if $r(N + e) = r(N)$; otherwise e is *independent of* N .

A *base* is a maximal independent submatroid of M , i. e. a matroid B in M such that $n(B) = 0$, while $B \subset N$, $B \neq N$ implies $n(N) > 0$. See also Theorem 7. A *base complement* $A = M - B$ is the complement in M of a base B . A *circuit* is a minimal dependent matroid, i. e. a matroid P such that $n(P) > 0$, while $N \subset P$, $N \neq P$ implies $n(N) = 0$.³

THEOREM 1. *N is independent if and only if it is contained in a base, or, if and only if it contains no circuit.*

³ Compare G, Theorem 9.

THEOREM 2. *A circuit is a minimal submatroid contained in no base, i. e. containing at least one element from each base complement. A base is a maximal submatroid containing no circuit. A base complement is a minimal submatroid containing at least one element from each circuit.*

The above facts follow at once from the definitions. Note the reciprocal relationship between circuits and base complements. Note also that the definitions of independence and of being a circuit depend only on the given subset, while the property of being a base depends on the relationship of the subset to M .

3. Properties of rank. Our object here is to prove Theorem 3. The following definition will be useful:

$$(3.1) \quad \Delta(M, N) = r(M + N) - r(M).$$

LEMMA 3. $\Delta(M + e_2, e_1) \leq \Delta(M, e_1)$.

Suppose first $r(M + e_1) = r(M) + 1$; then $r(M + e_1 + e_2) = r(M) + k$, $k = 1$ or 2 . If $k = 2$, then $r(M + e_2) = r(M) + 1$, on account of (R_2) , and the inequality holds; if $k = 1$, $r(M + e_2) = r(M) + l$, $l = 0$ or 1 , and it holds again. If $r(M + e_2) = r(M) + 1$, the same reasoning applies. If finally $r(M + e_1) = r(M + e_2) = r(M)$, the inequality follows from (R_3) .

LEMMA 4. $\Delta(M + N, e) \leq \Delta(M, e)$.

If $N = e_1 + \cdots + e_p$, the last lemma gives

$$\Delta(M + N, e) \leq \Delta(M + e_1 + \cdots + e_{p-1}, e) \leq \cdots \leq \Delta(M, e).$$

THEOREM 3. $\Delta(M + N_2, N_1) \leq \Delta(M, N_1)$, or,

$$(3.2) \quad r(M + N_1 + N_2) \leq r(M + N_1) + r(M + N_2) - r(M).$$

This is true if N_1 contains but a single element. For the general case, we apply the last lemma and induction, setting $N_1 = N' + e$:

$$\begin{aligned} \Delta(M + N_2, N_1) &= \Delta(M + N_2 + e, N') + \Delta(M + N_2, e) \\ &\leq \Delta(M + e, N') + \Delta(M, e) = \Delta(M, N_1). \end{aligned}$$

(3.2) is evidently equivalent to:

$$(3.3) \quad r(M_1 + M_2) \leq r(M_1) + r(M_2) - r(M_1 M_2).$$

4. Deduction of (I_1) , (I_2) from (R_1) , (R_2) , (R_3) . The first postulate

on independent sets below obviously holds if (R_1) and (R_2) hold. To prove (I_2) , take N, N' as given there; then

$$r(N) = p, \quad r(N') = p + 1.$$

We must show that for some i , $\Delta(N, e'_i) = 1$. (Then e'_i does not lie in N .) If this is not so, then on using Lemma 4 we find

$$\begin{aligned} 1 &= r(N') - r(N) \leq \Delta(N, N') \\ &= \Delta(N, e'_1) + \Delta(N + e'_1, e'_2) + \cdots + \Delta(N + e'_1 + \cdots + e'_p, e'_{p+1}) \\ &\leq \Delta(N, e'_1) + \Delta(N, e'_2) + \cdots + \Delta(N, e'_{p+1}) = 0, \end{aligned}$$

a contradiction.

5. Deduction of (C_1) , (C_2) from (R_1) , (R_2) , (R_3) . We shall need here a theorem showing how the nullity (or rank) of a matroid may be determined when we know what circuits it contains.

LEMMA 5. *Each element of a circuit is dependent on the rest of the circuit.*

If e is an element of the circuit P , then $n(P) = 1$, $n(P - e) = 0$; hence $r(P) = \rho(P) - 1 = \rho(P - e) = r(P - e)$.

LEMMA 6. *If e is dependent on P_1 but on no proper subset of P_1 , then $P = P_1 + e$ is a circuit.*

As $\Delta(P_1, e) = 0$, $r(P) = r(P_1) \leq \rho(P_1) < \rho(P)$, $n(P) > 0$, and P contains a circuit P' . If P' does not contain e , take e' in P' ; then

$$\Delta(P_1 - e', e') \leq \Delta(P' - e', e') = 0,$$

hence $r(P_1 - e') = r(P_1)$, and

$$\begin{aligned} \Delta(P_1 - e', e) &= r(P_1 - e' + e) - r(P_1 - e') \\ &\leq r(P_1 + e) - r(P_1) = \Delta(P_1, e) = 0, \end{aligned}$$

and e is dependent on the proper subset $P_1 - e'$ of P_1 , a contradiction. Therefore P' contains e . As P' is a circuit, e is dependent on the rest of P' ; hence $P' = P$.

THEOREM 4. *If e is not in N , there is a circuit in $N + e$ which contains e if and only if e is dependent on N .*

Suppose $P_1 + e = P$ is a circuit, $P_1 \subset N$. Then

$$\Delta(N, e) \leq \Delta(P_1, e) = 0,$$

and e is dependent on N . Suppose, conversely, $\Delta(N, e) = 0$. Let P_1 be a smallest subset of N on which e is dependent; then by the last lemma, $P = P_1 + e$ is a circuit. (It may be that $P = e$.)

THEOREM 5. *If N is formed element by element, then $n(N)$ is just the number of times that adding an element increases the number of circuits present.*

Say $N = e_1 + \cdots + e_p$. Then if O is the null set,

$$r(N) = \Delta(O, e_1) + \Delta(e_1, e_2) + \cdots + \Delta(e_1 + \cdots + e_{p-1}, e_p).$$

Each $\Delta(e_1 + \cdots + e_{i-1}, e_i) = 0$ or 1 , and $= 0$ if and only if e_i is dependent on $e_1 + \cdots + e_{i-1}$, i.e. if and only if there is a circuit in $e_1 + \cdots + e_i$ containing e_i . The number of terms is $p = \rho(N)$, and the theorem follows.

We turn now to the proof of (C_1) and (C_2) . The first is obvious. To prove the second, take P_1, P_2, e_1, e_2 as given. As

$$\Delta(P_1 - e_2, e_2) = \Delta(P_2 - e_1, e_1) = 0,$$

we have

$$\Delta(P_1 + P_2 - e_2, e_2) = \Delta(P_1 + P_2 - e_1 - e_2, e_1) = 0.$$

These equations give

$$r(P_1 + P_2 - e_1 - e_2) = r(P_1 + P_2 - e_2) = r(P_1 + P_2).$$

Using (R_2) gives

$$r(P_1 + P_2 - e_1) = r(P_1 + P_2 - e_1 - e_2);$$

hence the required circuit P_3 exists, by Theorem 4.

6. Postulates for independent sets. Let M be a set of elements. Let any subset N of M be either "independent" or "dependent." Let the two following postulates be satisfied:

(I_1) *Any subset of an independent set is independent.*

(I_2) *If $N = e_1 + \cdots + e_p$ and $N' = e'_1 + \cdots + e'_{p+1}$ are independent, then for some i such that e'_i is not in N , $N + e'_i$ is independent.*

The resulting system is equivalent to a matroid, as we now show. Given any subset N of M , we let $r(N)$ be the number of elements in a largest independent subset of N . Obviously Postulates (R_1) and (R_2) are satisfied; we must prove (R_3) . Say

$$r(N + e_1) = r(N + e_2) = r(N) = r.$$

Then $r(N + e_1 + e_2) = r$ or $r + 1$. If it equals $r + 1$, there is an independent set $N' = e'_1 + \cdots + e'_{r+1}$ in $N + e_1 + e_2$. Let $N'' = e''_1 + \cdots + e''_r$ be an independent set in N . By (I_2) there is an i such that $N'' + e'_i$ is an independent set of $r + 1$ elements. But $N'' + e'_i$ lies in $N + e_1$ or in $N + e_2$, and hence $r(N + e_1)$ or $r(N + e_2) \geq r + 1$, a contradiction. Therefore $r(N + e_1 + e_2) = r$, as required.

We have shown how to deduce either set of postulates (R) or (I) from the other. Moreover the definitions of the rank and the independence or dependence of any subset of M agree under the two systems, and hence they are equivalent.

7. Postulates for bases. Let M be a set of elements, and let each subset either be or not be a "base." We assume

(B_1) *No proper subset of a base is a base.*

(B_2) *If B and B' are bases and e is an element of B , then for some element e' in B' , $B - e + e'$ is a base.*

We shall prove the equivalence of this system with the preceding one. We write here $e_1 e_2 \cdots$ instead of $e_1 + e_2 + \cdots$ for short.

THEOREM 6. *All bases contain the same number of elements.*

For suppose

$$\begin{aligned} B &= e_1 \cdots e_p e_{p+1} \cdots e_q e_{q+1} \cdots e_r, \\ B' &= e_1 \cdots e_p e'_{p+1} \cdots e'_q \end{aligned}$$

are bases, with exactly e_1, \cdots, e_p in common, and $r > q$. We might have $p = 0$. $q > p$, on account of (B_1) . By (B_2) , we can replace e_{p+1} in B by an element e' of B' , giving a base B_1 . $e' = e'_{i_1}$ is one of the elements e'_{p+1}, \cdots, e'_q , for otherwise B_1 would be a proper subset of B . Hence

$$B_1 = e_1 \cdots e_p e'_{i_1} e_{p+2} \cdots e_q e_{q+1} \cdots e_r.$$

If $q > p + 1$, we replace e_{p+2} in B_1 by an element e'_{i_2} of B' , giving a base B_2 . Continuing in this manner, we obtain finally the base

$$B_{q-p} = e_1 \cdots e_p e'_{p+1} \cdots e'_q e_{q+1} \cdots e_r.$$

But this contains B' as a proper subset, contradicting (B_1) .

We shall say a subset of M is independent if it is contained in a base. (I_1) obviously holds; we shall prove (I_2) . Let N, N' be independent sets in the bases B, B' . Say

$$\begin{aligned} B &= e_1 \cdots e_p e_{p+1} \cdots e_q e_{q+1} \cdots e_r e_{r+1} \cdots e_s, \\ B' &= e_1 \cdots e_p e'_{p+1} \cdots e'_q e'_{q+1} \cdots e'_r e_{r+1} \cdots e_s, \\ N &= e_1 \cdots e_p e_{p+1} \cdots e_q, \quad N' = e_1 \cdots e_p e'_{p+1} \cdots e'_q e'_{q+1}. \end{aligned}$$

Then N and N' have just e_1, \cdots, e_p in common, and B and B' have just these elements and e_{r+1}, \cdots, e_s in common. By (B_2) , there is an element e'_{i_1} of B' such that

$$B_1 = B - e_{q+1} + e'_{i_1}$$

is a base. (This element cannot be any of $e_1, \cdots, e_p, e_{r+1}, \cdots, e_s$, by (B_1) .) If i_1 is one of the numbers $p+1, p+2, \cdots, q+1$, then $N + e'_{i_1}$ is in a base B_1 , as required. Suppose not; then there is a base

$$B_2 = B_1 - e_{q+2} + e'_{i_2}$$

with $i_2 \neq i_1$. If $p+1 \leq i_2 \leq q+1$, $N + e'_{i_2}$ is in a base B_2 . If not, we find a base B_3 , etc. We can drop out each of the $r-q$ elements e_{q+1}, \cdots, e_r in turn; as there are only $r-q-1$ elements e'_i with $i > q+1$, we find at some point a base containing e_1, \cdots, e_q, e'_j with $p+1 \leq j \leq q+1$. Then e'_j is in N' , and $N + e'_j$ is in a base and is thus independent, as required.

The definitions of base and independent sets in the two systems (I) and (B) are easily seen to agree. Suppose (I_1) and (I_2) hold. (B_1) obviously holds; using (I_2) , we prove that all bases contain the same number of elements; (B_2) now follows at once from (I_2) . Hence the two systems are equivalent.

THEOREM 7. *B is a base in M if and only if*

$$r(B) = r(M), \quad n(B) = 0.$$

Evidently B is a base under the given conditions. To prove the converse, we note first that there exists a base with $r(M)$ elements, as $r(M)$ is the maximum number of independent elements in M (see § 6). By Theorem 6, all bases have this many elements, and the equations follow.

THEOREM 8. *If B is a base and N is independent, then for some N' in B , $N + N'$ is a base.*

This follows from repeated application of Postulate (I₂) and the last theorem.

8. Postulates for circuits. Let M be a set of elements, and let each subset either be or not be a "circuit." We assume:

(C₁) *No proper subset of a circuit is a circuit.*

(C₂) *If P_1 and P_2 are circuits, e_1 is in both P_1 and P_2 , and e_2 is in P_1 but not in P_2 , then there is a circuit P_3 in $P_1 + P_2$ containing e_2 but not e_1 .*

(C₂) may be phrased as follows: If the circuits P_1 and P_2 have the common element e , then $P_1 + P_2 - e$ is the union of a set of circuits.

We shall define the rank of any subset of M , and shall then show that the postulates for rank are satisfied. Let e_1, \dots, e_p be any ordered set of elements of M . Set $\Gamma_i = 0$ if there is a circuit in $e_1 + \dots + e_i$ containing e_i , and set $\Gamma_i = 1$ otherwise (compare Theorem 5). Let the "rank" of (e_1, \dots, e_p) be

$$r(e_1, \dots, e_p) = \sum_{i=1}^p \Gamma_i.$$

LEMMA 7. $r(e_1, \dots, e_{q-2}, e_{q-1}, e_q) = r(e_1, \dots, e_{q-2}, e_q, e_{q-1})$.

To prove this, let N be the ordered set e_1, \dots, e_{q-2} , and set

$$\begin{aligned} r(N) &= r, & r(N, e_{q-1}) &= r_1, & r(N, e_q) &= r_2, \\ r(N, e_{q-1}, e_q) &= r_{12}, & r(N, e_q, e_{q-1}) &= r_{21}. \end{aligned}$$

CASE 1. There is no circuit in $N + e_{q-1}$ containing e_{q-1} , and none in $N + e_q$ containing e_q . Then

$$r_1 = r_2 = r + 1.$$

If there is a circuit in $N + e_{q-1} + e_q$ containing e_{q-1} and e_q , then

$$r_{12} = r_1 = r_2 = r_{21};$$

otherwise,

$$r_{12} = r_1 + 1 = r_2 + 1 = r_{21}.$$

CASE 2. There is a circuit P_2 in $N + e_{q-1}$ containing e_{q-1} , and a circuit P_1 in $N + e_{q-1} + e_q$ containing e_{q-1} and e_q . Then, by (C₂), there is a circuit P_3 in $N + e_q$ containing e_q . Hence

$$r_{12} = r_1 = r = r_2 = r_{21}.$$

CASE 3. There is a circuit P_2 as above, but no circuit P_1 as above. If there is a circuit P_3 as above, the last set of equations hold. Otherwise,

$$r_{12} = r_1 + 1 = r + 1 = r_2 = r_{21}.$$

CASE 4. There is a circuit in $N + e_q$ containing e_q . This case overlaps the two preceding ones; the proof above applies here also.

LEMMA 8. *The rank of any subset N is independent of the ordering of the elements of N .*

We saw above that interchanging the last two elements of any subset does not alter the rank; hence, evidently, interchanging any two adjacent elements leaves the rank unchanged. Any ordering of M may be obtained from any other by a number of interchanges of adjacent elements; the rank remains unchanged at each step, proving the lemma.

Postulates (R_1) and (R_2) are obviously satisfied. To prove (R_3) , suppose $r(N + e_1) = r(N + e_2) = r(N)$. Then there is a circuit in $N + e_1$ containing e_1 and one in $N + e_2$ containing e_2 ; hence $r(N + e_1 + e_2) = r(N)$.

The definitions of rank and of circuits under the two systems (R) , (C) agree, and hence the systems are equivalent.

9. **Fundamental sets of circuits.** The circuits P_1, \dots, P_q of a matroid M form a *fundamental set of circuits* if $q = n(M)$ and the elements e_1, \dots, e_n of M can be ordered so that P_i contains e_{n-q+i} but no e_{n-q+j} ($j > i$). The set is *strict* if P_i contains e_{n-q+i} but no e_{n-q+j} ($0 < j < i$ or $j > i$). These sets may be called sets *with respect to* e_{n-q+1}, \dots, e_n .

THEOREM 9. *If $B = e_1 + \dots + e_{n-q}$ is a base in $M = e_1 + \dots + e_n$, then there is a strict fundamental set of circuits with respect to e_{n-q+1}, \dots, e_n ; these circuits are uniquely determined.*

As $r(B) = r(M)$, $\Delta(B, e_i) = 0$ ($i = n - q + 1, \dots, n$). Hence, by Theorem 4, there is a circuit P_i containing e_i and elements (possibly) of B . P_{n-q+1}, \dots, P_n is the required set. Suppose, for a given i , there were also a circuit $P'_i \neq P_i$. Then Postulate (C_2) applied to P_i and P'_i would give us a circuit P in B , which is impossible.

This theorem corresponds to the theorem that if a square submatrix N of a matrix M is non-singular, then N can be turned into the unit matrix by a linear transformation on the rows of M .

THEOREM 10. *If P_1, \dots, P_q form a fundamental set of circuits with*

respect to e_{n-q+1}, \dots, e_n , then there is a unique strict set P'_1, \dots, P'_q with respect to e_{n-q+1}, \dots, e_n .

Set $B = M - (e_{n-q+1} + \dots + e_n)$. The existence of P_1, \dots, P_q shows that $r(M) = r(M - e_n) = \dots = r(B)$. Hence $\rho(B) = n - q = r(M) = r(B)$, and B is a base, by Theorem 7. Theorem 9 now applies.

Note that a matroid is not uniquely determined by a fundamental set of circuits (but see the appendix). This is shown by the following two matroids, in each of which the first two circuits form a strict fundamental set:

M , with circuits 1234, 1256, 3456;

M' , with circuits 1234, 1256, 13456, 23456.

II. SEPARABILITY, DUAL MATROIDS.

10. Separable matroids. If $M = M_1 + M_2$, then $r(M) \leq r(M_1) + r(M_2)$, on account of (3.3). If it is possible to divide the elements of M into two groups, M_1 and M_2 , each containing at least one element, such that

$$(10.1) \quad r(M) = r(M_1) + r(M_2),$$

or, which is equivalent (as M_1 and M_2 have no common elements),

$$(10.2) \quad n(M) = n(M_1) + n(M_2),$$

we shall say M is *separable*; otherwise, M is *non-separable*.⁴ Any single element forms a non-separable matroid. Any maximal non-separable part of M is a *component* of M .⁵

THEOREM 11. *If*

$$\begin{aligned} M &= M_1 + M_2, & r(M) &= r(M_1) + r(M_2), \\ M'_1 &\subset M_1, & M'_2 &\subset M_2, & M' &= M'_1 + M'_2, \end{aligned}$$

then

$$r(M') = r(M'_1) + r(M'_2).$$

Set $M_1'' = M_1 - M'_1$, $M_2'' = M_2 - M'_2$. The relations (see Theorem 3)

$$\begin{aligned} r(M) &= \Delta(M_1 + M'_2, M_2'') + \Delta(M', M_1'') + r(M') \\ &\leq \Delta(M'_2, M_2'') + \Delta(M'_1, M_1'') + r(M') \\ &= r(M_2) - r(M'_2) + r(M_1) - r(M'_1) + r(M') \end{aligned}$$

⁴ Compare G, Theorem 15.

⁵ See G, § 4.

together with the fact that $r(M) = r(M_1) + r(M_2)$ show that $r(M') \geq r(M'_1) + r(M'_2)$ and hence $r(M') = r(M'_1) + r(M'_2)$.

THEOREM 12.⁶ If $M = M_1 + M_2$, $r(M) = r(M_1) + r(M_2)$, M' is non-separable, and $M' \subset M$, then either $M' \subset M_1$ or $M' \subset M_2$.

For suppose $M' = M'_1 + M'_2$, $M'_1 \subset M_1$, $M'_2 \subset M_2$, and M'_1 and M'_2 each contain an element. By the last theorem, $r(M') = r(M'_1) + r(M'_2)$, which cannot be.

THEOREM 13. If M_1 and M_2 are non-separable matroids with a common element e , then $M = M_1 + M_2$ is non-separable.

For suppose $M = M'_1 + M'_2$, $r(M) = r(M'_1) + r(M'_2)$. By the last theorem, $M_1 \subset M'_1$ or $M_1 \subset M'_2$, and $M_2 \subset M'_1$ or $M_2 \subset M'_2$; this shows that either M'_1 or M'_2 is void.

THEOREM 14. No two distinct components of M have common elements.

This is a consequence of the last theorem. From this follows:

THEOREM 15.⁷ Any matroid may be expressed as a sum of components in a unique manner.

THEOREM 16.⁸ A non-separable matroid M of nullity 1 is a circuit, and conversely.

If M_1 is a proper non-null subset of the non-separable matroid M , and $M_2 = M - M_1$, then $r(M) < r(M_1) + r(M_2)$. Hence

$$1 = n(M) > n(M_1) + n(M_2),$$

and $n(M_1) = 0$, proving that M is a circuit.

Conversely, if $M = M_1 + M_2$ is a circuit, and M_1 and M_2 each contain elements, then

$$\begin{aligned} r(M_1) + r(M_2) &= \rho(M_1) + \rho(M_2) - n(M_1) - n(M_2) \\ &= \rho(M) > r(M), \end{aligned}$$

showing that M is non-separable.

⁶ Compare G, Lemma, p. 344.

⁷ Compare G, Theorem 12.

⁸ Compare G, Theorem 10.

LEMMA 9. Let $M = M_1 + M_2$ be non-separable, and let M_1 and M_2 each contain elements but have no common elements. Then there is a circuit P in M containing elements of both M_1 and M_2 .

Suppose there were no such circuit. Say $M_2 = e_1 + \cdots + e_s$. Using Theorem 4, we see that

$$\Delta(M_1 + e_1 + \cdots + e_{i-1}, e_i) = \Delta(e_1 + \cdots + e_{i-1}, e_i) \quad (i = 1, \cdots, s),$$

and hence $r(M) = r(M_1) + r(M_2)$, a contradiction.

THEOREM 17.⁹ Any non-separable matroid M of nullity $n > 0$ can be built up in the following manner: Take a circuit M_1 ; add a set of elements which forms a circuit with one or more elements of M_1 , forming a non-separable matroid M_2 of nullity 2 (if $n(M) > 1$); repeat this process till we have $M_n = M$.

As $n > 0$, M contains a circuit M_1 . If $n > 1$, we use the preceding lemma $n - 1$ times. The matroid at each step is non-separable, by Theorems 16 and 13.

THEOREM 18.¹⁰ Let $M = M_1 + \cdots + M_p$, and let M_1, \cdots, M_p be non-separable. Then the following statements are equivalent:

- (1) M_1, \cdots, M_p are the components of M .
- (2) No two of the matroids M_1, \cdots, M_p have common elements, and there is no circuit in M containing elements of more than one of them.
- (3) $r(M) = r(M_1) + \cdots + r(M_p)$.

We cannot replace rank by nullity in (3); see G, p. 347.

(2) follows from (1) on application of Theorems 13 and 16.

To prove (1) from (2), take any M_i . If it is not a component of M , there is a larger non-separable submatroid M'_i of M containing it. By Lemma 9, there is a circuit P in M'_i containing elements of M_i and elements not in M_i ; P must contain elements of some other M_j , a contradiction.

Next we prove (3) from (1). If $p > 1$, M is separable; say $M = M'_1 + M'_2$, $r(M) = r(M'_1) + r(M'_2)$. By Theorem 12, each M_i is in either M'_1 or M'_2 ; hence M'_1 and M'_2 are each a sum of components of M . If one of these

⁹ See G, Theorem 19; also Whitney, "2-isomorphic graphs," *American Journal of Mathematics*, vol. 55 (1933), p. 247, footnote.

¹⁰ Compare G, Theorem 17.

contains more than one component, we separate it similarly, etc. (3) now follows easily.

Finally we prove (1) from (3). Let M' be a component of M , and suppose it has an element in M_i . As

$$r(M) = r(M_i) + \sum_{j \neq i} r(M_j),$$

M' is contained in M_i , by Theorem 12; as M_i is non-separable, $M' = M_i$.

THEOREM 19.¹¹ *The elements e_1 and e_2 are in the same component of M if and only if they are contained in a circuit P .*

If e_1 and e_2 are both in P , they are part of a non-separable matroid, which lies in a single component of M . Suppose now e_1 and e_2 are in the same component M_0 of M , and suppose there is no circuit containing them both. Let M_1 be e_1 plus all elements which are contained in a circuit containing e_1 . By Lemma 9, there is a subset M^* of $M_0 - M_1$ which forms with part of M_1 a circuit P_3 . P_3 does not contain e_1 . If e'_4 is an element of P_3 in M_1 , there is a circuit P_1 in M_1 containing e_1 and e'_4 . Let e_3 be an element of M^* . Then in $M_1 + M^*$ there are circuits P_1 and P_3 which contain e_1 and e_3 respectively, and have a common element.

Let M' be a smallest subset of M_0 which contains circuits P'_1 and P'_3 such that one contains e_1 , the other contains e_3 , and they have common elements. Then P'_1 and P'_3 are distinct, and $M' = P'_1 + P'_3$. Let e_4 be a common element. By Postulate (C_2) , there is a circuit P_1 in $M' - e_4$ containing e_1 , and a circuit P_3 in $M' - e_4$ containing e_3 . By the definition of M' , P_1 and P_3 have no common elements. By Postulate (C_1) , P_1 is not contained in P'_1 ; hence it contains an element e_5 of $M' - P'_1$. P_3 does not contain e_5 . As P_3 is not contained in P'_3 , it contains an element e_6 of P'_1 . But now P'_1 contains e_1 , P_3 contains e_3 , $P'_1 + P_3$ have a common element e_6 , and $P'_1 + P_3$ does not contain e_5 and is thus a proper subset of M' , a contradiction. This proves the theorem.

11. Dual matroids. Suppose there is a 1—1 correspondence between the elements of the matroids M and M' , such that if N is any submatroid of M and N' is the complement of the corresponding matroid of M' , then

$$(11.1) \quad r(N') = r(M') - n(N).$$

¹¹ Compare D. König, *Acta Litterarum ac Scientiarum Szeged*, vol. 6, pp: 155-179, 4. (p. 159). The present theorem shows that a "glied" is the same as a component.

We say then that M' is a *dual* of M .¹²

THEOREM 20. *If M' is a dual of M , then*

$$r(M') = n(M), \quad n(M') = r(M).$$

Set $N = M$; then $n(N) = n(M)$. In this case N' is the null matroid, and $r(N') = 0$. (11.1) now gives $r(M') = n(M)$. Also

$$n(M') = \rho(M') - r(M') = \rho(M) - n(M) = r(M).$$

THEOREM 21. *If M' is a dual of M , then M is a dual of M' .*

Take any N and corresponding N' as before. The equations

$$\begin{aligned} r(N') &= r(M') - n(N), & r(M') &= n(M), \\ \rho(N) + \rho(N') &= \rho(M) \end{aligned}$$

give

$$\begin{aligned} r(N) &= \rho(N) - n(N) = \rho(N) - [r(M') - r(N')] \\ &= \rho(N) - n(M) + [\rho(N') - n(N')] \\ &= \rho(M) - n(M) - n(N') = r(M) - n(N'), \end{aligned}$$

as required.

THEOREM 22. *Every matroid has a dual.*

This is in marked contrast to the case of graphs, for only a planar graph has a dual graph (see G, Theorem 29).

Let M' be a set of elements in 1 — correspondence with elements of M . If N' is any subset of M' , let N be the complement of the corresponding subset of M , and set $r(N') = n(M) - n(N)$. (R_1) , (R_2) , (R_3) are easily seen to hold in M' , as they hold in M ; hence M' is a matroid. Obviously $r(M') = n(M)$, and M' is a dual of M .

THEOREM 23. *M and M' are duals if and only if there is a 1 — 1 correspondence between their elements such that bases in one correspond to base complements in the other.*

Suppose first M and M' are duals. Let B be a base in either matroid, say in M , and let B' be the complement of the corresponding submatroid of the other matroid, M' . Then

¹² Compare G, § 8. Theorems 20, 21, 24, 25 correspond to Theorems 20, 21, 23, 25 in G. Note that *two duals of the same matroid are isomorphic*, that is, there is a 1 — 1 correspondence between their elements such that corresponding subsets have the same rank. Such a statement cannot be made about graphs. Compare H. Whitney, "2-isomorphic graphs," *American Journal of Mathematics*, vol. 55 (1933), pp. 245-254.

$$\begin{aligned}r(B') &= r(M') - n(B) = r(M'), \\n(B') &= r(M) - r(B) = 0,\end{aligned}$$

and B' is a base in M' , by Theorem 7.

Suppose, conversely, that bases in one correspond to base complements in the other. Let N be a submatroid of M and let N' be the complement of the corresponding submatroid of M' . There is a base B' in M' with $r(N')$ elements in N' , by Theorem 8. The complement in M of the submatroid corresponding to B' in M' is a base B in M with $\rho(N') - r(N') = n(N')$ elements in $M - N$, and hence with $r(M) - n(N')$ elements in N . This shows that

$$r(N) = r(M) - n(N') + k, \quad k \geq 0.$$

In a similar fashion we see that

$$r(N') = r(M') - n(N) + k', \quad k' \geq 0.$$

As B contains $r(M)$ elements and B' contains $r(M')$ elements, $r(M) + r(M') = \rho(M)$. Hence, adding the above equations,

$$\begin{aligned}k + k' &= r(N) + r(N') + n(N) + n(N') - r(M) - r(M') \\&= \rho(N) + \rho(N') - \rho(M) = 0.\end{aligned}$$

Hence $k = 0$, and the first equation above shows that M and M' are duals.

There are various other ways of stating conditions on certain submatroids of M and M' which will ensure these matroids being duals.¹³

THEOREM 24. *Let M_1, \dots, M_p and M'_1, \dots, M'_p be the components of M and M' respectively, and let M'_i be a dual of M_i ($i = 1, \dots, p$). Then M' is a dual of M .*

Let N be any submatroid of M , and let the parts of N in M_1, \dots, M_p be N_1, \dots, N_p . Let N'_i be the complement in M'_i of the submatroid corresponding to N_i ; then $N' = N'_1 + \dots + N'_p$ is the complement in M' of the submatroid corresponding to N . By Theorems 18 and 11 we have

$$r(N') = r(N'_1) + \dots + r(N'_p), \quad n(N) = n(N_1) + \dots + n(N_p).$$

Also

$$r(M') = r(M'_1) + \dots + r(M'_p), \quad r(N'_i) = r(M'_i) - n(N_i);$$

adding the last set of equations gives $r(N') = r(M') - n(N)$, as required.

¹³ See for instance a paper by the author "Planar graphs," *Fundamenta Mathematicae*, vol. 21 (1933), pp. 73-84, Theorem 2. Cut sets may of course be defined in terms of rank.

THEOREM 25. Let M and M' be duals, and let M_1, \dots, M_p be the components of M . Let M'_1, \dots, M'_p be the corresponding submatroids of M' . Then M'_1, \dots, M'_p are the components of M' , and M'_i is a dual of M_i ($i = 1, \dots, p$).

The complement in M of the submatroid corresponding to M'_i in M' is $\sum_{j \neq i} M_j$. Hence, as M and M' are duals and the M_j ($j \neq i$) are the components of $\sum_{j \neq i} M_j$ (see Theorem 18),

$$r(M'_i) = r(M') - n\left(\sum_{j \neq i} M_j\right) = r(M') - \sum_{j \neq i} n(M_j).$$

Adding gives

$$\begin{aligned} \sum_i r(M'_i) &= pr(M') - (p-1) \sum_j n(M_j) = pr(M') - (p-1)n(M) \\ &= pr(M') - (p-1)r(M') = r(M'). \end{aligned}$$

Therefore, by Theorem 12, each component of M' is contained in some M'_i . In the same way we see that each component of M is contained in a matroid corresponding to a component of M' ; hence the components of one matroid correspond exactly to the components of the other.

Let N_i be any submatroid of M_i , and let N' and N'_i be the complements in M' and M'_i of the submatroid corresponding to N_i . The equations

$$\begin{aligned} r(M') &= \sum_j r(M'_j), & r(N') &= r(N'_i) + \sum_{j \neq i} r(M'_j), \\ r(N') &= r(M') - n(N_i), \end{aligned}$$

give

$$r(N'_i) = r(M'_i) - n(N_i),$$

which shows that M'_i is a dual of M_i .

THEOREM 26. A dual of a non-separable matroid is non-separable.

This is a consequence of the last theorem.

III. MATRICES AND MATROIDS.

12. Matrices, matroids, and hyperplanes. Consider the matrix

$$M = \left\| \begin{array}{cccc} a_{11} & \cdots & a_{1n} \\ . & . & . & . \\ a_{m1} & \cdots & a_{mn} \end{array} \right\|;$$

be C_1, \dots, C_n . Any subset N of these columns forms a matrix has a rank, $r(N)$. If we consider the columns as we have a matroid M . The proof of this is simple if we take a matrix as the number of linearly independent columns are then obvious. To prove (R_3) , suppose $r(N + C_1) = r(N)$; then C_1 and C_2 can each be expressed as a linear combination of columns of N , and hence $r(N + C_1 + C_2) = r(N)$. The (R_1) and base carry over to matrices and agree with the definition of a base in M is a minimal set of columns in terms of columns of M may be expressed.

M geometrically in two different ways; the second is for our purposes:

Let E_m be Euclidean space of m dimensions. Corresponding to each column of M is a point X_i in E_m with coordinates a_{i1}, \dots, a_{im} . The set of columns of M is linearly independent if and only if the points X_{i_1}, \dots, X_{i_p} are linearly independent in E_m , i. e. if and only if $p + 1$ points determine a hyperplane in E_m of dimension p . A subset N of M corresponds to a minimal set of points X_{i_1}, \dots, X_{i_p} in E_m such that N lies in the hyperplane determined by $O, X_{i_1}, \dots, X_{i_p}$. Then N is a base of M .

Let E_n be Euclidean space of n dimensions. Let R_1, \dots, R_m be the rows of M . If Y_1, \dots, Y_m are the corresponding points of E_n : $Y_i = (a_{i1}, \dots, a_{in})$, then the points O, Y_1, \dots, Y_m determine a hyperplane $H = H(M)$, which we shall call the *hyperplane associated with M* . The dimension $d(H)$ of H is $r(M)$. Let $N = C_{i_1} + \dots + C_{i_p}$ be a subset of M , and let E' be the p -dimensional coordinate subspace of E_n containing the x_{i_1} and \dots and the x_{i_p} axes. The j -th row of N corresponds to the point Y'_j in E' with coordinates $(a_{ji_1}, \dots, a_{ji_p})$; this is just the projection of Y_j onto E' . If H' is the hyperplane in E' determined by the points O, Y'_1, \dots, Y'_m , then H' is exactly the projection of H onto E' , and

$$(12.1) \quad d(H') = r(N).$$

Let $N = (C_{i_1}, \dots, C_{i_p})$ be any subset of M , and let E', H' correspond to N . Then N is independent if and only if

$$d(H') = p,$$

and is a base if and only if

$$d(H') = d(H) = p.$$

THEOREM 27. *There is a unique matroid M associated with plane H through the origin in E_n .*

Let M contain the elements e_1, \dots, e_n , one corresponding to each element of E_n . Given any subset e_{i_1}, \dots, e_{i_r} , we let its rank be the dimension of the projection of H onto the corresponding coordinate space. It was seen above that if M is any matroid determining a plane H , then M is the unique matroid associated with M .

13. Orthogonal hyperplanes and dual matroids. The following theorem:

THEOREM 28. *Let H be a hyperplane through the origin in E_n of dimension r , and let H' be the orthogonal hyperplane to H of dimension $n - r$. Let M and M' be the associated matroids. M and M' are duals.*

We shall show that bases in one matroid correspond to bases in the other; Theorem 23 then applies. Let

$$M = \left\| \begin{matrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{r1} & \cdots & a_{rn} \end{matrix} \right\|, \quad M' = \left\| \begin{matrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n-r,1} & \cdots & b_{n-r,n} \end{matrix} \right\|$$

be matrices determining H and H' respectively. Say the first r columns form a base in M , i. e. the corresponding determinant A is $\neq 0$. As H and H' are orthogonal, we have for each i and j

$$a_{i1}b_{j1} + a_{i2}b_{j2} + \cdots + a_{in}b_{jn} = 0.$$

Keeping j fixed, we have a set of r linear equations in the b_{jk} . Transpose the last $n - r$ terms in each equation to the other side, and solve for b_{jk} . We find

$$b_{jk} = \frac{-1}{A} \sum_{l=r+1}^n b_{jl} \begin{vmatrix} a_{11} & \cdots & a_{1l} & \cdots & a_{1r} \\ \vdots & & \vdots & & \vdots \\ a_{r1} & \cdots & a_{rl} & \cdots & a_{rr} \end{vmatrix} = \sum_{l=r+1}^n c_{kl} b_{jl} \quad (k = 1, \dots, r).$$

This is true for each $j = 1, \dots, n - r$, and the c_{kl} are independent of j . Thus the k -th column of M' is expressed in terms of the last $n - r$ columns. As this is true for $k = 1, \dots, r$, the last $n - r$ columns form a base in M' , as required.

14. The circuit matrix of a given matrix. Consider the matrix M of § 12. Suppose the columns C_{i_1}, \dots, C_{i_p} form a circuit, i. e. the corresponding

elements of the corresponding matroid form a circuit. Then these columns are linearly dependent, and there are numbers b_1, \dots, b_n such that

$$(14.1) \quad \begin{aligned} a_{i1}b_1 + \dots + a_{in}b_n &= 0 & (i=1, \dots, m), \\ b_j &= 0 \quad (j \neq i_1, \dots, i_p), & b_j \neq 0 \quad (j = i_1, \dots, i_p). \end{aligned}$$

The b_j are all $\neq 0$ ($j = i_1, \dots, i_p$), for otherwise a proper subset of the columns would be dependent, contrary to the definition of a circuit. (They are uniquely determined except for a constant factor; see Lemma 11.) Suppose the circuits of \mathbf{M} are P_1, \dots, P_s . Then there are corresponding sets of numbers b_{i1}, \dots, b_{in} ($i=1, \dots, s$), forming a matrix

$$\mathbf{M}' = \begin{vmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{s1} & \dots & b_{sn} \end{vmatrix},$$

the *circuit matrix* of the matrix \mathbf{M} .

THEOREM 29. *Let P_1, \dots, P_q be a fundamental set of circuits in \mathbf{M} (see § 9). Then the corresponding rows of the circuit matrix \mathbf{M}' form a base for the rows of \mathbf{M}' . Hence $r(\mathbf{M}') = q = n(\mathbf{M})$.*

Suppose the columns of \mathbf{M} are ordered so that P_i contains C_{n-q+i} but no column C_{n-q+j} ($j > i$). Then if the corresponding row of \mathbf{M}' is $R'_i = (b_{i1}, \dots, b_{in})$, we have $b_{i, n-q+i} \neq 0$ and $b_{i, n-q+j} = 0$ ($j > i$). Hence the rows R'_1, \dots, R'_q of \mathbf{M}' are linearly independent, and $r(\mathbf{M}') \geq q$. Hence $r(\mathbf{M}') = n(\mathbf{M}) = q$, and each row of \mathbf{M}' may be expressed in terms of R'_1, \dots, R'_q .

THEOREM 30. *If \mathbf{M}' is the circuit matrix of \mathbf{M} and H', H are the corresponding hyperplanes, then H' is the hyperplane of maximum dimension orthogonal to H .*

This is a consequence of (14.1) and the last theorem.

THEOREM 31. *The matroids corresponding to a matrix and its circuit matrix are duals.*

This follows from the last theorem and Theorem 28.

15. On the structure of a circuit matrix. Let \mathbf{M} be any matroid, and \mathbf{M}' , its dual. If there exists a matrix \mathbf{M} corresponding to \mathbf{M} , it is perhaps most easily constructed by considering it as the circuit matrix of a matrix \mathbf{M}'

corresponding to M' . Let H and H' be the hyperplanes corresponding to M and M' . We shall say the set of numbers (a_1, \dots, a_n) is in $Z_{i_1 \dots i_p}$ if

$$a_j \neq 0 \quad (j = i_1, \dots, i_p), \quad a_j = 0 \quad (j \neq i_1, \dots, i_p).$$

If (a_1, \dots, a_n) is in H and in $Z_{i_1 \dots i_p}$, then the columns C_{i_1}, \dots, C_{i_p} of M' are dependent, evidently.

LEMMA 10. *Let (b_1, \dots, b_n) be a point of H . If it is in $Z_{i_1 \dots i_p}$, then the matroid $N' = e_{i_1} + \dots + e_{i_p}$ is the union of a set of circuits in M' .*

Here e_i in M' corresponds to C_i in M . We need merely show that for each i_s there is a circuit P in N' containing e_{i_s} . Let $k_1 = i_s, k_2, \dots, k_q$ be a minimal set of numbers from (i_1, \dots, i_p) containing i_s such that there is a point (c_1, \dots, c_n) of H in $Z_{k_1 \dots k_q}$; then $e_{k_1} + \dots + e_{k_q}$ is the required circuit. For if it were not a circuit, there would be a proper subset (l_1, \dots, l_r) of (k_1, \dots, k_q) and a point (d_1, \dots, d_n) of H in $Z_{l_1 \dots l_r}$. No $l_i = k_1$, on account of the minimal property of (k_1, \dots, k_q) . Say $l_1 = k_t$, and set

$$a_i = d_{k_t} c_i - c_{k_t} d_i \quad (i = 1, \dots, n).$$

Then (a_1, \dots, a_n) is in H and in $Z_{m_1 \dots m_u}$ with (m_1, \dots, m_u) a proper subset of (k_1, \dots, k_q) again a contradiction.

LEMMA 11. *If $P = e_{i_1} + \dots + e_{i_p}$ is a circuit of M' and (b_1, \dots, b_n) and (b'_1, \dots, b'_n) are in H and in $Z_{i_1 \dots i_p}$, then these two sets are proportional.*

For otherwise, (c_1, \dots, c_n) with $c_i = b'_i b_i - b_i b'_i$ would be a point of H in some $Z_{k_1 \dots k_q}$ with (k_1, \dots, k_q) a proper subset of (i_1, \dots, i_p) , and P would not be a circuit.

It is instructive to show directly that Postulate (C_2) holds for matrices: P_1 and P_2 are represented by rows (b_1, \dots, b_n) and (b'_1, \dots, b'_n) of M , lying in $Z_{12i_1 \dots i_p}$ and $Z_{1k_1 \dots k_q}$ respectively, where $k_1, \dots, k_q \neq 2$. Set $c_i = b'_i b_i - b_i b'_i$; then (c_1, \dots, c_n) is in H and in $Z_{2i_1 \dots i_p, l_1 \dots l_r}$, with (l_1, \dots, l_r) a subset of $(i_1, \dots, i_p, k_1, \dots, k_q)$; the existence of P_3 now follows from Lemma 10.

THEOREM 32. *Let M be the circuit matrix of M' . Let P_1, \dots, P_q form a strict fundamental set of circuits in M' with respect to e_{n-q+1}, \dots, e_n , and let the first q rows in M correspond to P_1, \dots, P_q . Let (i_1, \dots, i_s) be any set of numbers from $(1, \dots, q)$, let (j_1, \dots, j_s) be any set from $(1, \dots, n-q)$, and let (i'_1, \dots, i'_{q-s}) be the set complementary to (i_1, \dots, i_s) in $(1, \dots, q)$.*

Then the determinant D in \mathbf{M} with rows i_1, \dots, i_s and columns j_1, \dots, j_s equals zero if and only if the determinant D' with rows $1, \dots, q$ and columns $j_1, \dots, j_s, n-q+i'_1, \dots, n-q+i'_{q-s}$ equals zero, or, if and only if there exists a circuit P in \mathbf{M}' containing none of the columns $e_{j_1}, \dots, e_{j_s}, e_{n-q+i'_1}, \dots, e_{n-q+i'_{q-s}}$.

In the matrix of the last $q = r(\mathbf{M})$ columns of \mathbf{M} , the terms along the main diagonal and only those are $\neq 0$. If we expand D' by Laplace's expansion in terms of the columns $n-q+i'_1, \dots, n-q+i'_{q-s}$, we see at once that $D' = 0$ if and only if $D = 0$.

Suppose $D = 0$. Then there is a set of numbers $(\alpha_1, \dots, \alpha_q)$, not all zero, with $\alpha_i = 0$ ($i \neq i_1, \dots, i_s$), such that

$$b_k = \alpha_1 b_{1k} + \dots + \alpha_q b_{qk} = 0 \quad (k = j_1, \dots, j_s),$$

$(b_{i_1}, \dots, b_{i_n})$ being the i -th row of \mathbf{M} , $b_k = 0$ also for $k = n-q+i'_1, \dots, n-q+i'_{q-s}$, as each term is zero for such k . The point (b_1, \dots, b_n) is in H . Any circuit given by Lemma 10 is the required circuit P .

Suppose the circuit P exists. Then it is represented by a row (b_1, \dots, b_n) in \mathbf{M} . As the first q rows of \mathbf{M} are of rank $q = r(\mathbf{M})$, (b_1, \dots, b_n) can be expressed in terms of them; say $b_k = \sum \alpha_i b_{ik}$. As $b_k = 0$ ($k = n-q+i'_1, \dots, n-q+i'_{q-s}$), certainly $\alpha_k = 0$ ($k = i'_1, \dots, i'_{q-s}$). $D = 0$ now follows from the fact that $b_k = 0$ ($k = j_1, \dots, j_s$).

16. A matroid with no corresponding matrix.¹⁴ The matroid \mathbf{M}' has seven elements, which we name $1, \dots, 7$. The bases consist of all sets of three elements except

$$(16.1) \quad 124, 135, 167, 236, 257, 347, 456.$$

Defining rank in terms of bases, we have: Each set of k elements is of rank k if $k \leq 2$ and of rank 3 if $k \geq 4$; a set of three elements is of rank 2 if the set is in (16.1) and is of rank 3 otherwise. It is easy to see that the postulates for rank are satisfied. (R_3) in the case that N contains two elements is satisfied vacuously. For suppose $r(N + e_1) = r(N + e_2) = r(N) = 2$. Then $N + e_1$ and $N + e_2$ are both in (16.1); but any two of these sets have but a single element in common.

¹⁴ After the author had noted that \mathbf{M}' satisfies (C^*) and corresponds to no linear graph, and had discovered a matroid with nine elements corresponding to no matrix, Saunders MacLane found that \mathbf{M}' corresponds to no matrix, and is a well known example of a finite projective geometry (see O. Veblen and J. W. Young, *Projective Geometry*, pp. 3-5).

If there exists a matrix M' , corresponding to M' , then let M be its circuit matrix. 123 is a base in M' , and hence

$$(16.2) \quad 124, 135, 236, 1237$$

form a fundamental set of circuits in M' . Let R_1, R_2, R_3, R_4 be the corresponding rows of M . By multiplying in succession row 1, column 2, rows 2, 3, 4, and columns 4, 5, 6, 7 by suitable constants $\neq 0$, we bring M into the following form:

$$(16.3) \quad M = \left\| \begin{array}{ccc|cccc} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & a & 0 & 1 & 0 & 0 \\ 0 & 1 & b & 0 & 0 & 1 & 0 \\ 1 & c & d & 0 & 0 & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{array} \right\|;$$

a, b, c and d are $\neq 0$. We now apply Theorem 32 with

$$(i_1, \dots, i_s; j_1, \dots, j_s) = (1, 4; 1, 2), (2, 4; 1, 3), (3, 4; 2, 3),$$

i. e. using the circuits 347, 257, 167. This gives

$$\begin{vmatrix} 1 & 1 \\ 1 & c \end{vmatrix} = \begin{vmatrix} 1 & a \\ 1 & d \end{vmatrix} = \begin{vmatrix} 1 & b \\ c & d \end{vmatrix} = 0,$$

and hence $c = 1, a = d = b$. Using the circuit 456, with sets $(1, 2, 3; 1, 2, 3)$ gives $2a = 0, a = 0$, a contradiction.

In regard to this example, see the end of the paper.

APPENDIX.

MATRICES OF INTEGERS MOD 2.

We wish to characterize those matroids M corresponding to matrices M of integers mod 2,¹⁵ i. e. matrices whose elements are all 0 or 1, where rank etc. is defined mod 2. We shall consider linear combinations, *chains*:

$$(A.1) \quad \alpha_1 e_1 + \dots + \alpha_n e_n \quad (\alpha\text{'s integers mod } 2)$$

in the elements of M . The α 's may be taken as 0 or 1; (A.1) may then be interpreted as the submatroid N whose elements have the coefficient 1. Conversely, any $N \subset M$ may be written as a chain. Submatroids are added

¹⁵ See O. Veblen, "Analysis situs," 2nd ed., *American Mathematical Society Colloquium Publications*, Ch. I and Appendix 2.

(mod 2) by adding the corresponding chains (mod 2). For instance, $(e_1 + e_2) + (e_2 + e_3) \equiv e_1 + e_3 \pmod{2}$.

Any sum (mod 2) of circuits in M we shall call a *cycle* in M . N is the *true sum* of N_1, \dots, N_s if these latter have no common elements and $N = N_1 + \dots + N_s$. We consider matroids which satisfy the following postulate:

(C*) *Each cycle is a true sum of circuits.*

Postulate (C₂) is a consequence of (C*). For the cycle $P_1 + P_2$ is a submatroid containing e_2 but not e_1 ; The existence of P_3 now follows from (C*).

A simple example of a matroid not satisfying (C*) is given by the matroid M' at the end of § 9.

THEOREM 33. *A circuit is a minimal non-null cycle, and conversely.*

This is proved with the aid of Postulates (C₁) and (C*).

THEOREM 34. *Let P_1, \dots, P_q be a strict fundamental set of circuits in M with respect to e_{n-q+1}, \dots, e_n . Then there are exactly 2^q cycles in M , formed by taking all sums (mod 2) of P_1, \dots, P_q .*

First, each sum $P_{i_1} + \dots + P_{i_s} \pmod{2}$ is a cycle, containing $e_{n-q+i_1}, \dots, e_{n-q+i_s}$ and elements (perhaps) from $B = e_1, \dots, e_{n-q}$; obviously distinct sums give distinct cycles. Now let Q be any cycle in M ; say Q contains $e_{n-q+k_1}, \dots, e_{n-q+k_r}$ and elements (perhaps) from B . Set $Q' = P_{k_1} + \dots + P_{k_r}$; then $Q + Q'$ is a cycle containing elements from B alone. But B is a base (see the proof of Theorem 10), and hence contains no circuits. Consequently $Q + Q'$ is the null cycle, and $Q = Q'$.

THEOREM 35. *As soon as the circuits of a strict fundamental set are known, all the circuits may be determined.*

This is a consequence of the last two theorems. It is to be contrasted with the final remark of § 9.

Remark. The word "strict" may be omitted in the last two theorems.

THEOREM 36. *Let e_1, \dots, e_n be a set of elements, and let P_1, \dots, P_q be any subsets such that P_i contains e_{n-q+i} and possibly elements from e_1, \dots, e_{n-q} alone. Then there is a unique matroid M satisfying (C*), with P_1, \dots, P_q as a strict fundamental set of circuits.*

We form the 2^q cycles of Theorem 34. Those cycles which contain no other non-null cycle as a proper subset we call circuits; in particular, P_1, \dots, P_q are circuits. To prove (C*), let Q be a non-null cycle. If it is not a circuit, it contains a circuit P as a proper subset. Q and P are sums (mod 2) from P_1, \dots, P_q , hence the same is true of $Q - P$, and $Q - P$ is one of the 2^q cycles. If it is not a circuit, we again extract a circuit, etc.

This theorem furnishes a simple method of constructing all matroids satisfying (C*).

We turn now to the study of matrices of integers (mod 2)

$$M = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{vmatrix} \quad (\text{each } a_{ij} = 0 \text{ or } 1).$$

Any linear combination (mod 2) of the columns

$$(A.2) \quad \alpha_1 C_1 + \cdots + \alpha_n C_n \quad (\alpha\text{'s integers mod } 2)$$

is a set of numbers $(\sum \alpha_i a_{i1}, \dots, \sum \alpha_i a_{mi})$, which we call a *chain* (mod 2) in M . As before, we may take each coefficient as 0 or 1, and we may consider a chain merely as a submatrix of M . The chain is a *cycle* if each of the corresponding numbers is $\equiv 0$ (mod 2). The columns C_{i_1}, \dots, C_{i_p} are *independent* (mod 2) if there exists no set of integers $\alpha_1, \dots, \alpha_n$ not all $\equiv 0$ (mod 2), with $\alpha_i = 0$ ($i \neq i_1, \dots, i_p$), such that $\sum \alpha_i C_i$ is a cycle, i. e. if no non-null subset of C_{i_1}, \dots, C_{i_p} is a cycle. Using this definition, the terms base, circuit, rank, nullity etc. (mod 2) can be defined as in Part I.

Let M be a set of elements e_1, \dots, e_n corresponding to C_1, \dots, C_n in M , and let $e_{i_1} + \cdots + e_{i_p}$ be a circuit in M if and only if C_{i_1}, \dots, C_{i_p} is a circuit in M . We shall show that M is a matroid satisfying (C*) and the definitions of cycle in M and M agree.

We show first that each circuit is a cycle in M . If C_{i_1}, \dots, C_{i_p} is a circuit, then these columns are dependent; hence $\sum \alpha_i C_i$ is a cycle, with $\alpha_i = 0$ ($i \neq i_1, \dots, i_p$). Moreover $\alpha_i = 1$ ($i = i_1, \dots, i_p$), for otherwise a proper subset of C_{i_1}, \dots, C_{i_p} would be dependent. Hence $C_{i_1} + \cdots + C_{i_p}$ is a cycle. Next, any sum (mod 2) of circuits is a cycle, evidently. Next we prove (C*). Suppose $Q = C_{i_1} + \cdots + C_{i_p}$ is a cycle. Let (k_1, \dots, k_q) be a minimal subset of (i_1, \dots, i_p) such that $P = C_{k_1} + \cdots + C_{k_q}$ is a cycle; then P is a circuit. $Q - P$ is a cycle; from it we extract a circuit, just as above, etc. It follows from (C*) that the definitions of cycle in M and M agree. Theorems 33, 34 and 35 now apply to M also.

We are now ready to prove the final theorem:

THEOREM 37. *Let M be any matroid satisfying (C*). Suppose $\rho(M) = n$, and $e_1 + \cdots + e_{n-q}$ is a base. Then if M_1 is any matrix of integers (mod 2) with $n - q$ columns which are independent (mod 2), columns C_{n-q+1}, \cdots, C_n can be adjoined in a unique manner to M_1 , forming a matrix M of which the corresponding matroid is M .*

Let P_1, \cdots, P_q be a strict fundamental set of circuits in M with respect to e_{n-q+1}, \cdots, e_n (Theorem 9). Say $P_1 = e_{i_1} + \cdots + e_{i_p} + e_{n-q+1}$. Set $C_{n-q+1} \equiv C_{i_1} + \cdots + C_{i_p}$ (mod 2); this determines C_{n-q+1} as a column of 0's and 1's so that $P'_1 = C_{i_1} + \cdots + C_{i_p} + C_{n-q+1}$ is a circuit. (P'_1 is a cycle; (C*) shows that it is a single circuit, as $C_1 + \cdots + C_{n-q}$ contains no circuit.) C_{n-q+1} evidently must be chosen in this manner. We choose the remaining columns of M similarly. Let M' be the matroid corresponding to M . Then P'_1, \cdots, P'_q is a strict set of circuits in M' . These same sets form a strict set in M ; hence, by Theorem 35, the circuits in M' correspond to those in M . Consequently $M' = M$, completing the proof.

We end by noting that the matroid M' of § 16 satisfies Postulate (C*) but corresponds to no linear graph. For letting 123 be a base and (16.2) a fundamental set of circuits and determining the matroid as in Theorem 36, we come out with exactly M' . A corresponding matrix of integers mod 2 is constructed from (16.3) with $a = b = c = d = 1$; we interchange rows and columns in the left-hand portion, leave out the last row and column of the right-hand portion, and interchange these two parts. (The relation $2a = 0$ is of course true mod 2.)

On the other hand, it is easily seen that if the element 7 is left out, there is a corresponding graph, which must be of the following sort: It has four vertices a, b, c, d , and the arcs corresponding to the elements 1, \cdots , 6 are

$$ab, ac, ad, bc, bd, cd.$$

There is no way of adding the required seventh arc.

The problem of characterizing linear graphs from this point of view is the same as that of characterizing matroids which correspond to matrices (mod 2) with exactly two ones in each column.

ON THE ASYMPTOTIC DISTRIBUTION OF THE REMAINDER TERM OF THE PRIME-NUMBER THEOREM.

By AUREL WINTNER.

The result of the present note is to the effect that the Riemann hypothesis is equivalent not only with the best possible order of the remainder term of the prime-number theorem but—on a proper scale—also with a generalized almost-periodic behavior of this remainder term. This means that the *formally trigonometrical* development of the remainder term is a *Fourier* development, which implies, in particular, the existence of an asymptotic distribution function.

Let ρ_1, ρ_2, \dots be the sequence of distinct zeros of $\zeta(s)$ in the upper half-plane, so that

$$(1) \quad \rho_k = 1/2 + i\gamma_k, \quad \gamma_{k+1} > \gamma_k > 0 \quad (k = 1, 2, \dots)$$

by assumption. Let n_k denote the multiplicity of the zero ρ_k and let

$$(2) \quad \begin{aligned} \phi_0(x) &\equiv 0, \quad \phi_m(x) = x^{-1/2}(A_m + \bar{A}_m), \text{ where } x > 1, \\ A_m &= \sum_{k=1}^m n_k x^{\rho_k} / \rho_k, \end{aligned} \quad (m = 1, 2, \dots).$$

It is known¹ that

$$(3) \quad \phi(x) = \lim_{m \rightarrow \infty} \phi_m(x)$$

exists and that on placing, as usual,

$$\psi(x) = \sum_{p^i \leq x} \log p$$

the "explicit formula" of the prime-number theory may be written as

$$(4) \quad x - \psi(x) = x^{1/2} \phi(x) + \log[2\pi(1 - x^{-2})^{1/2}],$$

where $x \neq p^n$; at the discontinuity points, $x = p^n$, of $\psi(x)$ one has to replace $\psi(x)$ by the arithmetical mean of $\psi(x + 0)$ and $\psi(x - 0)$. The rôle of (1) for the distribution of the prime numbers is² that of implying for the remainder term (4) of the prime-number theorem $\psi(x) \sim x$ the appraisal $x^{1/2}O(x^\epsilon)$ for any $\epsilon > 0$, and even the appraisal

¹ Cf. E. Landau, *Vorlesungen über Zahlentheorie*, Leipzig, 1927, Theorem 452.

² *Ibid.*, Theorem 453.

$$(5) \quad \phi(x) = O(\log^2 x).$$

Finally, (1) is equivalent³ also with

$$(6) \quad (\log \omega)^{-1} \int_2^\omega (1 - \psi(x)/x)^2 dx \rightarrow \sum_{k=1}^{\infty} 2n_k^2 / |\rho_k|^2, \quad \omega \rightarrow \infty,$$

a relation which may be written, according to (4), in the form

$$(6a) \quad (\log \omega)^{-1} \int_2^\omega x^{-2} \{x^{1/2} \phi(x) + \log[2\pi(1-x^{-2})^{1/2}]\}^2 dx \\ \rightarrow 2 \sum_{k=1}^{\infty} n_k^2 / |\rho_k|^2, \quad \omega \rightarrow \infty.$$

Now not only (6a) holds but also

$$(7) \quad (\log \omega)^{-1} \int_2^\omega x^{-2} \{x^{1/2} [\phi(x) - \phi_{m-1}(x)] + \log[2\pi(1-x^{-2})^{1/2}]\}^2 dx \\ \rightarrow 2 \sum_{k=m}^{\infty} n_k^2 / |\rho_k|^2; \quad (m = 1, 2, \dots) \quad (\omega \rightarrow \infty).$$

If $m = 1$, then (7) reduces to (6a) or (6) in virtue of (2). While for $m \neq 1$ the relation (7) is not clear from (6a), (3) and (2), a glance at the proof of (6) shows that the proof of (7) needs but a repetition of the proof of (6a), so that the proof of (7) will be omitted.

On denoting the integrand of (7) by

$$(7a) \quad x^{-1} \{\phi(x) - \phi_{m-1}(x)\}^2 + D_m(x),$$

it follows from (1), (2) and (5) that

$$D_m(x) = 2x^{-3/2} [\phi(x) - \phi_{m-1}(x)] \log[2\pi(1-x^{-2})^{1/2}] + x^{-2} \log^2[2\pi(1-x^{-2})^{1/2}] \\ = 2x^{-3/2} [O(\log^2 x) + O(1)] O(1) + x^{-2} O(1)^2 = O(x^{-5/4});$$

hence

$$\int_2^\omega D_m(x) dx = \int_2^\omega O(x^{-5/4}) dx = O(1) = o(\log \omega).$$

Thus it is clear from (7), (7a) that, for every fixed m ,

$$(\log \omega)^{-1} \int_2^\omega x^{-1} [\phi(x) - \phi_{m-1}(x)]^2 dx \rightarrow 2 \sum_{k=m}^{\infty} n_k^2 / |\rho_k|^2, \quad \omega \rightarrow \infty,$$

i. e.,

$$T^{-1} \int_1^T [\phi(e^x) - \phi_{m-1}(e^x)]^2 dx \rightarrow 2 \sum_{k=m}^{\infty} n_k^2 / |\rho_k|^2, \quad T \rightarrow \infty.$$

³ *Ibid.*, Theorems 476 and 477. This result is due to Cramér.

Hence, on placing

$$(8) \quad M\{g\} = \lim_{T \rightarrow \infty} \int_1^T g(x) dx / T,$$

one has

$$(9) \quad M\{(f - s_{m-1})^2\} = 2 \sum_{k=m}^{\infty} n_k^2 / |\rho_k|^2,$$

where

$$s_m(x) = \phi_m(e^x), \quad f(x) = \phi(e^x),$$

so that

$$(10) \quad s_m(x) = \sum_{k=1}^m n_k (e^{i\gamma_k x} / \rho_k + e^{-i\gamma_k x} / \bar{\rho}_k)$$

and

$$(11) \quad f(x) = \sum_{k=1}^{\infty} n_k (e^{i\gamma_k x} / \rho_k + e^{-i\gamma_k x} / \bar{\rho}_k)$$

in virtue of (1), (2) and (3). The relations (4) and (5) take the form

$$(11a) \quad e^x - \psi(e^x) = e^{x/2} f(x) + \log 2\pi + O(e^{-2x}), \quad f(x) = O(x^2).$$

Now, from (9),

$$(12) \quad M\{(f - s_m)^2\} \rightarrow 0, \quad m \rightarrow \infty.$$

Due to (10) and (11), the relation (12) might be expressed by saying that the trigonometrical series (11) not only is convergent but that it is the Fourier series of the function which it represents, i. e., that

$$(12a) \quad f(x) \sim \sum_{k=1}^{\infty} n_k (e^{i\gamma_k x} / \rho_k + e^{-i\gamma_k x} / \bar{\rho}_k),$$

the equivalence sign \sim being understood in the Besicovitch sense.⁴ It must, however, be mentioned that the averaging process (8) operates not in the symmetric range $[-T, T]$ but only in the upper half of it; to the lower half of this range there corresponds the range $0 < x < 1$ of $\psi(x) \equiv 0$, where the behavior of the series $\phi(x)$ is just as intricate as in the range $1 < x < \infty$.

Since, however, (11) is a pure sine series, the formal difficulty just mentioned may be avoided by *defining* the function $f(x)$ for $-\infty < x < -1$ by $f(x) = -f(-x)$ and for $-1 \leq x \leq 1$ arbitrarily. Then $T^{-1} \int_1^T$ in (8) may be replaced by $(2T)^{-1} \int_{-T}^T$, so that the function $f(x)$ occurring in the fundamental formula (11a) belongs to the Besicovitch class B^2 in virtue of (12). As a consequence of this fact, it follows from the Besicovitch theory

⁴ A. S. Besicovitch, *Almost periodic functions*, Cambridge, 1932, Chap. II.

that the coefficients ⁵ n_k/ρ_k , $n_k/\bar{\rho}_k$ of the series (11) may be represented in the Fourier-Bohr manner as averages. It is clear from the definition of $f(x)$ for $x < -1$ that these expressions of the coefficients hold also if $M\{\}$ is understood in the sense (8). Unfortunately, nothing is known about the diophantine nature ⁶ of the frequencies $\pm \gamma_n$ of the Fourier expansion (12 a); in particular, it is not known in what manner the frequencies are generated by a basis of linearly independent numbers.

If $g(x)$ is a real-valued measurable function defined for $1 < x < \infty$, and if, for a given $T > 1$ and a given real number ξ , one denotes by $[g(x) < \xi; T]$ the set of those points x of the interval $1 < x < T$ at which $g(x) < \xi$, the function $g(x)$ is said to possess an asymptotic distribution function $\sigma = \sigma(\xi)$ if at every continuity point ξ of this σ the relation

$$T^{-1} \text{meas } [g(x) < \xi; T] \rightarrow \sigma(\xi), \quad (T \rightarrow \infty)$$

holds, and

$$\sigma(-\infty) = 0, \quad \sigma(+\infty) = 1.$$

Thus $\sigma(\xi)$ is monotone and not everywhere constant. The limit defining $\sigma(\xi)$ is zero for every ξ if, as x increases indefinitely, $g(x)$ is very often very large; hence in such a case there does not exist an asymptotic distribution function. Now $s_m(x)$ is, according to (10), real-valued and almost-periodic in the Bohr sense and has therefore ⁷ an asymptotic distribution function. Hence it follows ⁸ from (12) that $f(x)$ also possesses an asymptotic distribution function. This fact expresses a certain amount of regularity in the fluctuations of $f(x)$ and implies, in particular, that $|f(x)|$ cannot be very often very large. On the other hand, ⁹

$$(11 \text{ b}) \quad f(x) = \Omega_{\pm}(\log \log x), \quad x \rightarrow \infty,$$

so that neither $f(x)$ nor $-f(x)$ is less than a positive constant. While the

⁵ It is not known if $n_k = 1$ for every k .

⁶ For a property of the numbers γ_n which is, however, not of an arithmetical nature, cf. pp. 101-102 of this volume.

⁷ A. Wintner, "Diophantische Approximationen und Hermitesche Matrizen," I, *Mathematische Zeitschrift*, vol. 30 (1929), pp. 310-312. In the following year, Jessen, and Bohr and Jessen, also proved the existence of asymptotic distribution functions. Cf. also the programmatic address of Bohr in the *Proceedings of the 5th Scandinavian Congress* (1922). For the recent development of the distribution theory, cf. B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, July, 1935.

⁸ Cf. B. Jessen and A. Wintner, *loc. cit.*, Theorem 24.

⁹ Cf. E. Landau, *op. cit.*, Theorem 472, or H. Bohr, *loc. cit.* (11 b) is due to Littlewood.

asymptotic distribution function cannot be everywhere constant, it seems to be difficult to decide whether it is nowhere constant¹⁰ or whether $f(x)$ "dislikes" some regions $a < f(x) < b$. The answer to this question might depend¹¹ on the one mentioned at the end of the previous paragraph.

The relation (12 a) may be expressed also in terms of the Dirichlet series

$$F(s) = \sum_{k=1}^{\infty} n_k / \gamma_k e^{-\gamma_k s}, \quad \Re s > 0.$$

It is known¹² that (11 b) depends on the behavior of

$$\Im F(s) \text{ as } \Re s \rightarrow +0.$$

Now $\Im F(s)$ on the boundary line $\Re s = 0$ not only is a convergent trigonometrical series but is also the Fourier series of the function which it represents. In fact, the number of zeros of $\xi(s)$ in the strip $0 < \Im s < T$ is $O(T \log T)$, even if one counts every zero according to its multiplicity. This implies the convergence of the series

$$\sum_{k=1}^{\infty} n_k^2 / |\gamma_k \rho_k|.$$

Hence it follows from (11) that the series representing the sum of $f(x)$ and $2\Im F(ix)$ is a trigonometrical series which possesses a convergent majorant uniformly for all x and is therefore almost-periodic in the sense of Bohr. Accordingly the trigonometrical series

$$-\sum_{k=1}^{\infty} n_k / \gamma_k \sin \gamma_k x,$$

which defines the function $\Im F(ix)$, belongs to this function as its Fourier series (B^2). In particular, $\Im F(ix)$ is a function of class B^2 and has therefore an asymptotic distribution function.

THE JOHNS HOPKINS UNIVERSITY.

¹⁰ Cf. A. Wintner, "Remarks on the Ergodic Theorem of Birkhoff," *Proceedings of the National Academy of Sciences*, vol. 18 (1932), p. 251.

¹¹ Cf., in this connection, B. Jessen and A. Wintner, *loc. cit.*

¹² Cf. E. Landau, *op. cit.*, Theorem 470.

ON THE EXACT VALUE OF THE BOUND FOR THE REGULARITY OF SOLUTIONS OF ORDINARY DIFFERENTIAL EQUATIONS.

By AUREL WINTNER.

Let $f(z, w)$ be regular-analytic and bounded in the four-dimensional domain $|z| < a, |w| < b$, and let $w = w(z)$ be the solution of $dw/dz = f(z, w)$ which vanishes at $z = 0$. Let M denote the least upper bound of $|f(z, w)|$ in the domain $|z| < a, |w| < b$. It is known that there exists a bound $\Gamma = \Gamma(a, b, M)$ which is independent of the particular choice of $f(z, w)$ and is such that $w(z)$ is regular-analytic in the circle $|z| < \Gamma$. In fact, the method of successive approximations yields the estimate

$$(1) \quad \Gamma(a, b, M) \geq \min(a, b/M).$$

The necessity of the limitation $|z| < a$ is obvious from the case where $f(z, w)$ is independent of w and has singularities on the circle $|z| = a$. On the other hand, the necessity of the limitation $|z| < b/M$ is not evident. In fact, the latter limitation is introduced into the proof of (1) only for a somewhat artificial reason,—in order to assure the possibility of successive substitutions into f .

It turns out, however, that the trivial, and *a priori* artificial, appraisal (1) cannot be improved, i. e., that the value of the best bound $\Gamma(a, b, M)$ is precisely $\min(a, b/M)$. This situation seems to be unexpected insofar as efforts have been made¹ to improve the lower estimate (1) of $\Gamma(a, b, M)$. In reality, these efforts succeeded only by imposing additional restrictions on $f(z, w)$. Such a restriction is that $f(z, w)$ satisfies a uniform Lipschitz condition in the open domain $|z| < a, |w| < b$; and the corresponding improved estimate of the regularity radius of $w(z)$ depends* not only on a, b, M but also on the Lipschitz constant. A proof of the upper estimate

$$(2) \quad \Gamma(a, b, M) \leq \min(a, b/M),$$

which clears up this situation, runs as follows.

If $a \leq b/M$ then (2) is obvious from the case where $f(z, w)$ is independent of w . In order to prove that (2) holds also when $a > b/M$, it is clearly sufficient to show that for any *given* pair of numbers b, M and for any

¹ Cf. P. Painlevé, *Encyklopaedie der Mathematischen Wissenschaften*, vol. 1, Part I, p. 194 and p. 200.

given number r , where $r > b/M$, there exists a function f which is independent of z and possesses the following properties:

- (i) $f(w)$ is regular-analytic and bounded in the circle $|w| < b$;
- (ii) the least upper bound of $|f(w)|$ in this circle is M ;
- (iii) the function $w = w(z)$ for which $dw/dz = f(w)$ and $w(0) = 0$ has in the circle $|z| < r$ a singularity.

Now the function

$$(3) \quad f(w) = M[(1 + w/b)/2]^{1/n}$$

satisfies all these conditions if n is sufficiently large, larger than a number depending on r . In fact, the solution $w(z)$ belonging to (3) is

$$w(z) = b[(1 + z/C_n)^{n/(n-1)} - 1]$$

where

$$C_n = (1 - 1/n)^{-1} 2^{1/n} b/M.$$

Thus $z = -C_n$ is a singular point of $w(z)$ and tends to $z = -b/M$ when $n \rightarrow +\infty$. This proves (iii), while (i) and (ii) are satisfied by (3) for any n .

THE JOHNS HOPKINS UNIVERSITY.

ON SYMMETRIC BERNOULLI CONVOLUTIONS.

By RICHARD KERSHNER and AUREL WINTNER.

A class of symmetric Bernoulli convolutions which are regular analytic in the whole plane or in a strip containing the real axis or which possess, at least, a high degree of smoothness along the real axis, has recently been considered by one of the authors.¹ The present note deals mainly with the other extreme case, where the convolution possesses but a very low degree of smoothness. One class of convolutions which will be considered includes, for instance, the well known Cantor function; and other functions which have been treated in the literature also occur. These and some other examples have been collected in a joint paper of B. Jessen and one of the present authors.² The present note attempts a systematic treatment of a type of these symmetric convolutions with a low degree of smoothness. Apart from the theory of infinite convolutions, the functions to be considered are of interest from the point of view of the theory of real functions. The dominating feature of some of the convolutions in question is the homogeneous character of their spectra and a corresponding homogeneity of the mapping involved. In particular, one is lead to absolutely continuous convolutions which might be termed length-preserving with respect to a nowhere dense set of positive measure. It turns out that Bernoulli convolutions of this type are identical with the functions $\phi(x)$ considered by Hausdorff³ in connection with his fractional dimension theory. While Hausdorff is mainly interested in the case where the Lebesgue measure, which will be denoted by $\mu(E)$, is zero, his results hold for the case of a positive Lebesgue measure also, a case with which the present paper is mainly concerned. A class of Bernoulli convolutions which might be termed complementary to the case of the Hausdorff functions $\phi(x)$ also is considered.

Let $\beta(x)$ denote the symmetric Bernoulli distribution of standard de-

¹ A. Wintner, "On analytic convolutions of Bernoulli distributions," *American Journal of Mathematics*, vol. 56 (1934), pp. 659-663; "On symmetric Bernoulli convolutions," *Bulletin of the American Mathematical Society*, vol. 41 (1935), pp. 137-138.

² B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta-function," *Transactions of the American Mathematical Society*, vol. 37 (1935), § 6, Theorem 11.

³ F. Hausdorff, "Dimension und äusseres Mass," *Mathematische Annalen*, vol. 79 (1919), pp. 157-179.

viation 1 so that $\beta(x)$ is 0, $1/2$ or 1 according as x is on the left, in the interior or on the right of the interval $-1 < x < +1$. Thus $\beta(x/a)$, where $a > 0$, also is a symmetric Bernoulli distribution function; the distribution function

$$(1) \quad \frac{1}{2}(1 + \operatorname{sign} x),$$

which belongs to $a = +0$, will not be considered as a Bernoulli distribution function. The infinite Bernoulli convolution

$$(2) \quad \sigma(x) = \beta(x/a_1) * \beta(x/a_2) * \cdots$$

is convergent if and only if

$$(3) \quad \sum_{n=1}^{\infty} a_n^2 < +\infty,$$

and the convergence of (2) implies its absolute convergence.⁴ It will always be supposed that (3) is satisfied. The function (2) always is continuous.⁵ Further,⁶ if $\sigma(x)$ is not absolutely continuous, it is singular, i. e. such that $\sigma'(x) = 0$ almost everywhere. The spectrum S of $\sigma(x)$, defined as the set of points x in the vicinity of which σ is not constant, consists⁷ of those points x which are representable in the form

$$(4) \quad \sum_{n=1}^{\infty} \pm a_n,$$

where the signs depend on n in an arbitrary way, the only restriction being that the series be convergent. Hence S is a bounded set or the whole real axis according as the condition

$$(5) \quad \sum_{n=1}^{\infty} a_n < +\infty$$

is or is not satisfied. Examples show⁸ that σ may be singular or absolutely continuous whether (5) is satisfied or not, so that all four possibilities actually occur. The set S is always perfect, since the set of points in the vicinity of which a continuous function is not constant is either perfect or empty. On denoting by ρ_n the infinite convolution

$$(6) \quad \rho_n(x) = \beta(x/a_{n+1}) * \beta(x/a_{n+2}) * \cdots,$$

so that ρ_n tends, as $n \rightarrow +\infty$, to the distribution function (1), either all functions

⁴ B. Jessen and A. Wintner, *loc. cit.*

⁵ *Ibid.*

⁶ *Ibid.*

⁷ *Ibid.*

⁸ *Ibid.*, examples 1, 3, 5, and 6.

$$(7) \quad \rho_0 = \sigma, \rho_1, \rho_2, \dots$$

are singular or all are absolutely continuous. For if $\rho_n(x)$ is singular, then the derivative of

$$\rho_{n-1}(x) = \rho_n(x) * \beta(x/a_n) = \frac{1}{2}[\rho_n(x + a_n) + \rho_n(x - a_n)]$$

is zero almost everywhere; and if $\rho_n(x)$ is absolutely continuous, then so is $\rho_{n-1}(x)$, since absolute continuity cannot be lost by the convolution process.⁹ It may be mentioned that if S_n denotes the spectrum of ρ_n , then either all sets

$$(8) \quad S_0 = S, S_1, S_2, \dots$$

are nowhere dense or none are nowhere dense. For if S_n contains an interval, then so does each of the sets¹⁰ $S_n - a_n$ and $S_n + a_n$, the logical sum of which is S_{n-1} in virtue of (4); further, if S_n contains an interval, then so does S_{n+1} . For if I be an interval in S_n , then $I - a_{n+1}$ either has an interval in common with the perfect set S_{n+1} or contains a subinterval $J - a_{n+1}$ which does not contain any point of S_{n+1} . In the latter case $J + a_{n+1}$ is contained in S_{n+1} by the definition of S_n in terms of S_{n+1} , so that in either case S_{n+1} contains an interval.

From now on it will be supposed that S is bounded, i. e., that (5) is satisfied, so that one may introduce the remainders

$$(9) \quad r_n = \sum_{m=n+1}^{\infty} a_m \quad (n = 0, 1, 2, \dots).$$

The following theorem will now be proven:

If

$$(10) \quad a_n > \sum_{m=n+1}^{\infty} a_m = r_n$$

⁹ In fact, absolute continuity, in the case of a distribution function $\psi(x)$, means that there exists for every $\epsilon > 0$ a $\delta = \delta(\epsilon)$ such that

$$\sum_{k=1}^{\infty} |\psi(x'_k) - \psi(x''_k)| < \epsilon \quad \text{whenever} \quad \sum_{k=1}^{\infty} |x'_k - x''_k| < \delta.$$

Since the latter inequality implies that $\sum_{k=1}^{\infty} |(x'_k - y) - (x''_k - y)| < \delta$ for any y ,

it also implies that $\sum_{k=1}^{\infty} |\psi(x'_k - y) - \psi(x''_k - y)| < \epsilon$ for any y . Hence it implies that

$$\sum_{k=1}^{\infty} \left| \int_{-\infty}^{+\infty} \psi(x'_k - y) d\omega(y) - \int_{-\infty}^{+\infty} \psi(x''_k - y) d\omega(y) \right| \leq \epsilon \int_{-\infty}^{+\infty} d\omega(y) = \epsilon$$

for any distribution function ω or that $\psi * \omega$ is absolutely continuous for any distribution function ω .

¹⁰ By $B + c$ is meant the set of points representable in the form $x + c$, where x is a number contained in B .

for every n , then the spectrum S of the infinite Bernoulli convolution (2) is nowhere dense and has the measure

$$(11) \quad \mu(S) = 2 \sum_{n=1}^{\infty} a_n - \sum_{n=1}^{\infty} 2^n (a_n - \sum_{m=n+1}^{\infty} a_m) = 2 \lim_{n \rightarrow \infty} 2^n r_n,$$

which may or may not be zero. If $\mu(S) > 0$ and X denotes the interval $[-\infty, x]$, then

$$(12) \quad \sigma(x) = \mu(SX)/\mu(S),$$

i. e., the measure of the σ -image of an x -interval is proportional to the portion of the nowhere dense spectrum contained in the interval. In particular, $\sigma(x)$ is singular or absolutely continuous according as $\mu(S) = 0$ or $\mu(S) > 0$. If it is absolutely continuous, then its density is bounded¹¹ since σ satisfies a uniform Lipschitz condition.

The fact that $\mu(S) > 0$ implies absolute continuity is of interest since if (10) is not satisfied, then $\mu(S) > 0$ is necessary but not sufficient for the absolute continuity of $\sigma(x)$; cf. (23).^{11a} That in the case (10) the condition $\mu(S) > 0$ is sufficient for absolute continuity is clear from the relation (12), since (12) implies the uniform Lipschitz condition

$$|\sigma(x') - \sigma(x'')| \leq C |x' - x''|,$$

the best value of C being $1/\mu(S)$. The second representation of $\mu(S)$ given in (11) shows that $\mu(S) = 0$ or $\mu(S) > 0$ according as $r_n = o(2^{-n})$ is or is not satisfied, and that (10) implies the existence of $\lim 2^n r_n$.

It is seen from (9) that (10) may be written in the form

$$(13) \quad r_n > 2r_{n+1} \quad (n = 0, 1, 2, \dots).$$

This clearly implies the possibility of a successive construction of open sub-intervals J of the closed interval $[-r_0, r_0]$ as follows. Let J_{12} denote the open interval which is symmetric with respect to the mid-point of $[-r_0, r_0]$ and is of length $2(r_0 - 2r_1)$. From each of the two closed intervals K_1, K_2 which constitute $[-r_0, r_0] - J_{12}$ one may remove an open interval of length $2(r_1 - 2r_2)$ and having the mid-point of K_i , where $i = 1, 2$, as mid-point. This is possible since K_1 and K_2 are each of length $4r_1 > 2(r_1 - 2r_2)$. Let J_{14} and J_{34} denote the open intervals thus removed and let J_{14} be the one

¹¹ Up to a set of measure zero.

^{11a} For another example of this type cf. A. Denjoy, "Sur quelques points de la théorie des fonctions," *Comptes Rendus*, vol. 194 (1932), pp. 44-46, and H. Minkowski, *Gesammelte Abhandlungen*, vol. 2 (1911), pp. 50-51 and fig. 7.

which is to the left of J_{12} . From each of the four closed intervals which constitute

$$[-r_0, r_0] - J_{12} - J_{14} - J_{34}$$

one may remove open intervals $J_{13}, J_{33}, J_{53}, J_{73}$ in the same symmetric manner as J_{14} and J_{34} have been removed from K_1 and K_2 , it being understood that each of the four intervals J_{k3} is of length $2(r_2 - 2r_3)$ and that J_{k3} is on the left of J_{h3} if $k < h$. On continuing this process one obtains for every n

$$(14) \quad \begin{array}{l} 2^n \text{ intervals } J_{k2^{n+1}} \text{ of length } 2(r_n - 2r_{n+1}), \\ \text{where } k = 1, 3, 5, \dots, 2^{n+1} - 1 \end{array} \quad (n = 0, 1, 2, \dots).$$

It is convenient to write the double subscript of J_{pq} as a fraction by placing $J_{pq} = J_{p/q}$ so that J_t is defined for every number t of the form

$$(15) \quad t = \sum_{j=1}^m b_j / 2^j, \text{ where } b_j = 0 \text{ or } b_j = 1,$$

i. e., for every number of the interval $0 < t < 1$ having a finite dyadic development. J_t and J_u are, by their successive construction, disjoint if $t \neq u$. Now it is easy to verify¹² from the definition of $\beta(x/a_n)$ and from that of the convolution operator $*$ that

$$(16) \quad \sigma(x) \equiv t \text{ if } x \text{ is in } J_t.$$

Since the points (15) lie dense in the interval $0 < t < 1$ and since the distribution function $\sigma(x)$ is everywhere continuous, it follows that every subinterval of $[-r_0, r_0]$ contains a J_t . Consequently, the set

$$(17) \quad [-r_0, r_0] - \sum_t J_t,$$

where t runs through all values (15), is nowhere dense and consists of the cluster points of the endpoints of the open intervals J_t . Hence it is clear from (16) and from the definition of the spectrum S that the set (17) is contained in S . Since S is a subset of $[-r_0, r_0]$ in virtue of (4), the set (17) is precisely S . Accordingly, J_t and J_u being disjoint if $t \neq u$,

$$\mu(S) = \mu([-r_0, r_0]) - \mu(\sum_t J_t) = 2r_0 - \sum_t \mu(J_t),$$

so that

$$\mu(S) = 2r_0 - \sum_{n=0}^{\infty} 2^n (r_n - 2r_{n+1})$$

in virtue of (14). On comparing this with (9) one obtains (11).

¹² Cf. B. Jessen and A. Wintner, *loc. cit.*

It must now be shown that in the case $\mu(S) > 0$ the relation (12) holds.¹³ Since $\sigma(x)$ is non-decreasing and continuous, it is sufficient to verify (12) for a dense set of points x . Now $[-r_0, r_0]$ contains S , so that (12) is trivial if x is not in $[-r_0, r_0]$. Since $\sum_t J_t$ is dense in $[-r_0, r_0]$, it follows that it is sufficient to verify (12) for the points x of a J_t . Let x be in J_t and let $t = k/2^m$. Then the $2^m - 1$ intervals J_t , where $t = j/2^n$ and $j = 1, 3, 5, \dots, 2^{n-1}$; $n = 1, 2, 3, \dots, m$, decompose S into 2^m congruent parts, each of which has the measure $2^{-m}\mu(S)$ since the intervals J_t have been removed symmetrically. Since there are, among the 2^m congruent parts of measure $2^{-m}\mu(S)$, exactly k on the left of the point x , one has

$$\mu(SX) = k2^{-m}\mu(S).$$

This proves (12) since $t = k/2^m$, and $\sigma(x) = t$ by (16).

The Hausdorff theory¹⁴ of λ -measure and its further development by Besicovitch¹⁵ allow, of course, an analysis of the case $\mu(S) = 0$ also.

As an illustration of the theorem, let

$$(18) \quad a_n = Aa^n + Bb^n, \text{ where } 0 < a < b < 1, A > 0, B \geq 0.$$

It is easily verified that (10) is satisfied if and only if

$$(19) \quad b \leq 1/2$$

and that (11) gives $\mu(S) = 0$ or $\mu(S) = 2B$, according as $b < 1/2$ or $b = 1/2$. Thus if

$$(20) \quad a_n = B(1/2)^n + Aa^n, \text{ where } 0 < a < 1/2, B > 0, A > 0,$$

then (2) is absolutely continuous with a nowhere dense set of positive measure as spectrum and is represented by the formula (12).

The infinite convolution (2) belonging to the sequence (18) in the case $A = 1, B = 0$ will be denoted by $\sigma_a(x)$, so that

$$(21) \quad \sigma_a(x) = \beta(x/a) * \beta(x/a^2) * \beta(x/a^3) * \dots \quad (0 < a < 1).$$

Since (19) takes, in the case $B = 0$, the form $a < 1/2$ in virtue of $a < b$, the function $\sigma_a(x)$ is singular with a spectrum S of zero measure if $a < 1/2$. In particular, $\sigma_{1/3}(x)$ is the usual Cantor function considered by Lebesgue.

¹³ Cf. F. Hausdorff, *loc. cit.*, § 11.

¹⁴ F. Hausdorff, *loc. cit.*, §§ 10-12.

¹⁵ A. S. Besicovitch, "On linear sets of points of fractional dimension," *Mathematische Annalen*, vol. 101 (1929), pp. 161-193.

On the other hand, $\sigma_{1/2}(x)$ is the distribution function of the "Abrundungsfehler," i. e.,

$$\sigma_{1/2}(x) = (x+1)/2 \text{ if } -1 \leq x \leq 1$$

and $S = [-1, 1]$, so that $\sigma_{1/2}(x)$ is absolutely continuous with a bounded density. This example shows that on replacing $>$ in (10) by \geq , the spectrum S of (2) may become an interval. Since the sequence $\{(1/2)^n\}$ consists of the two sequences $\{(1/4)^n\}$ and $\{(1/4)^n/2\}$, it is clear that

$$(22 \text{ a}) \quad \sigma_{1/4}(x) * \sigma_{1/4}(2x) = \sigma_{1/2}(x);$$

this relation is an instance of the fact that the convolution of two singular Bernoulli convolutions may be absolutely continuous.¹⁶ On the other hand,

$$(22 \text{ b}) \quad \sigma_{1/4}(x) * \sigma_{1/4}(x)$$

is singular. This is shown in the same way¹⁷ as for $\sigma_{1/3}(x) * \sigma_{1/3}(x)$. The interest of the latter example lies in the fact that although it is singular, the spectrum is an interval, as will follow from (23). Besides, (23) will show that the spectrum of $\sigma_a(x)$ is an interval not only in the limiting case $a = 1/2$ but in the case $1/2 < a < 1$ as well.

If $>$ in (10) be replaced by \leq , then S becomes connected:

$$(23) \quad S = [-r_0, r_0] \text{ if } a_n \leq r_n \quad (n = 1, 2, \dots).$$

This is easily seen from (4) and (9) by an obvious extension of the proof of Riemann's theorem according to which

$$0 < a_n \rightarrow 0 \quad \text{and} \quad \sum_{n=1}^{\infty} a_n = +\infty$$

imply that every real number is representable in the form (4). The examples

$$a_n = (1/2)^n \text{ and } a_{2n} = a_{2n+1} = (1/3)^n$$

show that in the case (23) both absolutely continuous and singular convolutions (2) are possible.

If (10), i. e. $a_n > r_n$, is satisfied not for every n but only for sufficiently large values of n , then S is still nowhere dense since S_n in (8) is then nowhere dense if n is sufficiently large. If $a_n \leq r_n$ is satisfied not for every n but only for sufficiently large n , then S_n is, according to (23), an interval if n is

¹⁶ Cf. P. Lévy, "Sur les séries dont les termes sont des variables éventuelles indépendantes," *Studia Mathematica*, vol. 3 (1931), p. 153.

¹⁷ Cf. B. Jessen and A. Wintner, *loc. cit.*, example 2.

sufficiently large, so that $S = S_0$ consists of a finite number (≥ 1) of intervals in virtue of (4). The following example shows that S may consist of an arbitrarily large number of disjoint intervals.

For a given $\alpha > 1$, let $\sigma^\alpha(x)$ denote the Bernoulli convolution belonging to $a_n = n^{-\alpha}$, so that

$$(24) \quad \sigma^\alpha(x) = \beta(x/1^\alpha) * \beta(x/2^\alpha) * \beta(x/3^\alpha) * \cdots,$$

and let S^α denote the spectrum of (24). Since $a_n = n^{-\alpha}$ satisfies $a_n \leq r_n$ for sufficiently large n if α is fixed, the spectrum S^α consists of $N = N_\alpha$ disjoint intervals. It is easy to see that, as $\alpha \rightarrow +\infty$, the number of intervals increases indefinitely while S^α shrinks to the set consisting of the pair of points $x = \pm 1$. In this connection it is interesting to mention that¹⁸ $\sigma^\alpha(x)$ possesses, for every fixed α , derivatives of arbitrarily high order for every x , although $\sigma^\alpha(x)$ cannot, of course, be analytic at the end points of the N_α intervals which constitute S^α . It is not known whether $\sigma^\alpha(x)$ is or is not analytic in the interior of these intervals. Since every n may be written uniquely in the form $n = 2^k(2m+1)$, it is clear from (24) that¹⁹

$$(25) \quad \sigma^\alpha(x) = \sigma_c(x/1^\alpha) * \sigma_c(x/3^\alpha) * \sigma_c(x/5^\alpha) * \cdots, \text{ where } c = (1/2)^\alpha < 1/2,$$

so that $\sigma_c(x)$ is singular with a spectrum of zero measure. It may be mentioned that if α is near enough to 1, then $N_\alpha = 1$, and that $S^\alpha \rightarrow [-\infty, \infty]$ as $\alpha \rightarrow 1+0$.

THE JOHNS HOPKINS UNIVERSITY.

¹⁸ A. Wintner, *loc. cit.*

¹⁹ Cf. P. Lévy, *loc. cit.*, p. 154.

ON UNIFORM CONVERGENCE.

By J. W. THEODORE SUCKAC.

Introduction. While in Classical Analysis it is always explicitly presupposed that a convergent sequence of functions is uniformly convergent, it was first observed by Egoroff¹ that a very strong type of approximate uniform convergence is automatically present in every convergent sequence of measurable functions.

The theorem of Egoroff has led to many investigations, in particular by F. Riesz² who has applied the theorem in a very interesting manner to the Lebesgue theory.³

It is the purpose of this paper to investigate the phenomenon of uniform convergence in a general way so as to obtain a better understanding of the situation in the case of measurable functions.

I. *Uniform Convergence.* Suppose that the sequence $f_n(x)$ ⁴ is defined and convergent on a set S . If the sequence is uniformly convergent on some subset of S , that subset is said to have the character U (and is designated by the letter U).

We introduce a class of sets, Ψ , namely the totality of subsets of S having the character U .

The class Ψ is neither the null set nor does it contain only the null set, since any subset of S with a finite number of elements is a subset of character U . Moreover, the addition of a finite number of points of S to an element of Ψ yields another element in Ψ . Hence, except in the trivial case when the sequence is uniformly convergent on S there is no largest subset in the class Ψ .

¹ D. Th. Egoroff, "Sur les suites des fonctions mesurables," *Comptes Rendus*, Paris, vol. 152 (1910), pp. 244-246.

² F. Riesz, (i) "Sur l'intégrale de Lebesgue," *Acta Mathematica*, vol. 42 (1920), pp. 191-205; (ii) "Sur le théorème de M. Egoroff et sur les opérations fonctionnelles linéaires," *Acta Litt. Sci. Szeged*, vol. 1 (1922), pp. 18-25; (iii) "Elementarer Beweis des Egoroffschen Satzes," *Monatsheften für Mathematik und Physik*, vol. 25 (1928), pp. 243-248.

³ F. Riesz, footnote ² (i), above; *loc. cit.*, pp. 196-205.

⁴ Unless otherwise stated all functions are entirely unrestricted.

In the special case when S is itself an element of Ψ it is the largest element and the class Ψ is coincident with the totality of subsets of S . However, even when this is not the case it is possible by a very simple process⁵ to determine the class Ψ . This process is fundamental in the paper.

Let $S(k, \nu)$ be the subset of S on which the inequality

$$|f_n(x) - f_m(x)| \leq 1/k$$

holds for every $n, m \geq \nu$.

Take any sequence of positive integers

$$(\nu) : \nu_1, \nu_2, \nu_3, \dots$$

Finally, consider

$$U(\nu) = \prod_{k=1}^{\infty} S(k, \nu_k).$$

THEOREM. *The sequence $f_n(x)$ is uniformly convergent on $U(\nu)$, and conversely, if the sequence $f_n(x)$ is uniformly convergent on a subset S^* of S , then there exists a sequence (ν) such that $U(\nu)$ contains S^* . In other words the totality of sets $U(\nu)$ together with their subsets is the class Ψ .*

Proof. To prove the first part, take any $\epsilon > 0$ and pick k so that $1/k \leq \epsilon$. Now every x in $U(\nu)$ is also in $S(k, \nu_k)$, and so for every x in $U(\nu)$ it is true that $|f_n(x) - f_m(x)| \leq 1/k \leq \epsilon$ for $n, m \geq \nu_k$. Hence the sequence is uniformly convergent on $U(\nu)$.

Conversely, if $f_n(x)$ is uniformly convergent on S^* then given any positive integer k there exists a ν_k such that for all $n, m \geq \nu_k$ it is true that $|f_n(x) - f_m(x)| \leq 1/k$ for every x in S^* . Thus a sequence $(\nu) : \nu_1, \nu_2, \nu_3, \dots$ has been defined. S^* is contained in $S(k, \nu_k)$ for every value of k and therefore S^* is contained in $\prod_{k=1}^{\infty} S(k, \nu_k) = U(\nu)$.

1.2. Now that the class Ψ has been determined it is interesting to examine its elements. We know that every finite subset of S is contained in Ψ . Under certain conditions we are assured that at least one infinite subset of S is a member of Ψ .

THEOREM. *If S is non-denumerable, then there exists a denumerably infinite subset on which the convergence is uniform.*

Proof. There is an α_1 such that $S(1, \alpha_1)$ is non-denumerable. This is

⁵ Though not explicitly stated by Riesz, the process used by him is essentially the same. Footnote ² (iii), p. 549, *loc. cit.*, p. 244.

so since $S = \sum_{\nu=1}^{\infty} S(1, \nu)$ and hence if every $S(1, \nu)$ were denumerable it would follow that S had this property, which is contrary to hypothesis.

Pick x_1 from $S(1, \alpha_1)$.

Consider the sequence now as defined only on $S(1, \alpha_1)$ which is non-denumerable. There is an α_2 such that when the construction of 1.1 is considered with respect to $S(1, \alpha_1)$, then $S(2, \alpha_2: 2)$ is non-denumerable. ($S(2, \alpha_2: 2)$ is the set of points of $S(1, \alpha_1)$ where $|f_n(x) - f_m(x)| \leq \frac{1}{2}$ for $n, m \geq \alpha_2$).

Pick $x_2 \neq x_1$ from $S(2, \alpha_2: 2)$.

Moreover, there is a β_2 such that for all $n, m \geq \beta_2$

$$|f_n(x_i) - f_m(x_i)| \leq \frac{1}{2} \quad (i = 1).$$

.

Consider the sequence as defined only on $S(k-1, \alpha_{k-1}: k-1)$ which is non-denumerable. There is an α_k such that when the construction of 1.1 is considered with respect to $S(k-1, \alpha_{k-1}: k-1)$ then $S(k, \alpha_k: k)$ is non-denumerable. ($S(k, \alpha_k: k)$ is the set of points of $S(k-1, \alpha_{k-1}: k-1)$ where $|f_n(x) - f_m(x)| \leq 1/k$ for $n, m \geq \alpha_k$).

Pick $x_k \neq x_1, x_2, x_3, \dots, x_{k-1}$ from $S(k, \alpha_k: k)$.

Moreover, there is a β_k such that for $n, m \geq \beta_k$

$$|f_n(x_i) - f_m(x_i)| \leq 1/k \quad [i = 1, 2, 3, \dots, (k-1)].$$

Choose $\nu_k = \overline{\alpha_k, \beta_k}^6: \beta_1 = \alpha_1$.

Define $(\nu): \nu_1, \nu_2, \nu_3, \dots$

Then $U(\nu)$ contains the set x_1, x_2, x_3, \dots

$S(k, \alpha_k: k)$ certainly contains the set $x_k, x_{k+1}, x_{k+2}, \dots$, since $S(k, \alpha_k: k)$ contains $S(k-1, \alpha_{k-1}: k-1)$; while $x_1, x_2, x_3, \dots, x_{k-1}$ is included in $S(k, \beta_k)$. Furthermore, $S(k, \nu_k)$ contains both the sets $S(k, \alpha_k: k)$ and $S(k, \beta_k)$ and so it contains the denumerably infinite set x_1, x_2, x_3, \dots for every value of k .

This proves the theorem.

That the non-denumerability of S in the hypothesis of the above theorem is essential is shown by the following example.

$$\text{Define} \quad f_n(x) = \begin{cases} 0 & \text{on } 1, 1/2, 1/3, \dots, 1/n \\ 1 & \text{on } 1/n + 1, 1/n + 2, 1/n + 3, \dots \end{cases}$$

⁶ $\overline{a_1, a_2, a_3, \dots, a_n}$ is the largest of this set of values, $\underbrace{a_1, a_2, a_3, \dots, a_n}_{\text{the smallest.}}$

Now $f_n(x) \rightarrow 0$ on the set $\{1/n\}$, and yet on any infinite subset the convergence is not uniform.

This example and the above theorem may be combined in the statement of the following theorem:

A necessary and sufficient condition that each convergent sequence of functions defined on S have associated with it a denumerably infinite subset of S on which the convergence is uniform, is that S be non-denumerable.

1.3. The question arises as to whether or no this theorem is the best of its kind. We might perhaps always have a subset of character U which is non-denumerable. We are going to develop an example to show that this is not in general possible if we assume the hypothesis of the continuum, that $2^{\aleph_0} = \aleph_1$. In other words the above theorem is in a way final.

The example we shall develop is that of a sequence of functions converging to zero on the whole continuum and yet uniformly convergent on no non-denumerable subset.

Before going to the construction of the example we shall state two auxiliary theorems.

AUXILIARY THEOREM I. *If a null sequence⁷ of functions is uniformly convergent on a set S , then there is a null sequence of numbers which dominates⁸ the sequence of functions on S .*

This is merely a restatement of the definition of uniform convergence on S .

AUXILIARY THEOREM II. *Given any denumerable aggregate of null sequences (of numbers) there exists a null sequence dominated by none of the sequences of the aggregate.*

Proof. Suppose that the given sequences are:

$$\begin{aligned} a_{11}, a_{12}, a_{13}, \dots &\rightarrow 0 \\ a_{21}, a_{22}, a_{23}, \dots &\rightarrow 0 \\ a_{31}, a_{32}, a_{33}, \dots &\rightarrow 0 \\ \vdots & \end{aligned}$$

(i) Pick the first value of n , say $\nu(1)$ such that $|a_{1,\nu(1)}| < 1$.
Cancel the first $\nu(1)$ columns.

⁷ A null sequence is a sequence converging to zero.

⁸ The sequence a_n dominates the sequence $b_n(x)$ on S if there is a subscript n_0 such that for all $n > n_0$, $a_n \geq b_n(x)$, on S .

(ii) Pick the first remaining value of n , say $\nu(2)$ such that $|a_{1,\nu(2)}| < 1/2$.
 Cancel the next $\nu(2) - \nu(1)$ columns.
 Pick the first remaining value of n , say $\nu(3)$ such that $|a_{2,\nu(3)}| < 1/2$.
 Cancel the next $\nu(3) - \nu(2)$ columns.
 (iii) Pick the first remaining value of n , say $\nu(4)$ such that $|a_{1,\nu(4)}| < 1/3$.
 Cancel the next $\nu(4) - \nu(3)$ columns.
 Pick the first remaining value of n , say $\nu(5)$ such that $|a_{2,\nu(5)}| < 1/3$.
 Cancel the next $\nu(5) - \nu(4)$ columns.
 Pick the first remaining value of n , say $\nu(6)$ such that $|a_{3,\nu(6)}| < 1/3$.
 Cancel the next $\nu(6) - \nu(5)$ columns.

.

Now construct the sequence b_1, b_2, b_3, \dots in this wise:

$$\begin{aligned} b_1 &= b_2 = \dots = b_{\nu(1)} = 1 \\ b_{\nu(1)+1} &= \dots = b_{\nu(2)} = \dots = b_{\nu(3)} = 1/2 \\ b_{\nu(3)+1} &= \dots = b_{\nu(4)} = \dots = b_{\nu(5)} = \dots = b_{\nu(6)} = 1/3. \\ &\dots \end{aligned}$$

The sequence $\{b_n\}$ is a null sequence and it is evident that it is dominated by none of the given sequences.

We may now go to the actual *construction of the example*.

Normally order the following sets: (i) all real numbers, (ii) the totality of null sequences of numbers.

$$(i) \quad x_1, x_2, x_3, \dots, x_\lambda, \dots (\lambda < \Omega)$$

$$(ii) \quad A_1, A_2, A_3, \dots, A_\lambda, \dots (\lambda < \Omega)$$

where Ω is the ordinal which initiates the cardinal c .

The required sequence will be defined as an \aleph_c -valued function, $B(x)$.

Define $B(x_\lambda)$ to be the first sequence in (ii) which is not dominated by any A_ν for all $\nu \leq \lambda$. Suppose that the sequence is $A_{\mu(\lambda)}$.

It follows from Auxiliary Theorem II that there is a null sequence not dominated by A_ν for all $\nu \leq \lambda$; for, these null sequences constitute a denumerable aggregate, as λ is ordinally less than Ω the ordinal which initiates the cardinal c , and it is assumed that $c = 2^{\aleph_0} = \aleph_1$. Thus the definition of $B(x)$ is complete.

$B(x_\lambda)$ is possibly dominated by a null sequence, say A_ν , only for x 's with subscripts $\lambda < \nu$; for if $\nu \leq \lambda$, then by definition $B(x_\lambda)$ is not dominated by A_ν . In other words $B(x)$ is possibly dominated by A_ν only for the x 's in the set

$$x_1, x_2, x_3, \dots, x_\beta, \dots (\beta < \nu)$$

that is on at most a denumerable set.

Suppose that the sequence $A_{\mu(\lambda)}$ is

$$l_1, l_2, l_3, \dots, l_n, \dots$$

Define $f_n(x_\lambda) = l_n$.

Evidently $f_n(x) \rightarrow 0$. Moreover, the sequence $f_n(x)$ is uniformly convergent on no non-denumerable set; for if it were uniformly convergent on some non-denumerable set, then, by Auxiliary Theorem I, it would be dominated by a null sequence on this set, which would imply that $B(x)$ is dominated by a null sequence over a non-denumerable set, and this has been shown to be impossible.

1.4. Let us summarize these results in the following manner. In examining the elements of the class Ψ six possibilities present themselves.

- (i) The class Ψ is the null set.
- (ii) The class Ψ contains only the null set.
- (iii) The class Ψ contains only finite subsets of S .
- (iv) The class Ψ contains only denumerable subsets of S .
- (v) The class Ψ contains at least one non-denumerable subset of S .
- (vi) The class Ψ contains the set S .

The results of the preceding section show that:

I. If S is itself denumerable, then (i) and (ii) are impossible, but it is possible for (iii), (iv) and (vi) to occur.

II. If S is itself non-denumerable, then (i), (ii) and (iii) are impossible, but it is possible for (iv), (v) and (vi) to occur.

Hence it appears that in a special case where S is non-denumerable and no U is non-denumerable, the sequence is behaving with the utmost possible stubbornness with respect to the property of uniform convergence.

II. *Approximate uniform convergence.* So far in our investigation of the sets U in the class Ψ we have confined ourselves to their power. We now make a finer distinction and consider their measure.

A sequence $f_n(x)$ is said to be *approximately uniformly convergent* on S if for every positive ϵ there is a U such that $m_\epsilon(S - U) < \epsilon$.⁹ We would like to find conditions on the set S and the functions $f_n(x)$ which are sufficient to insure the above phenomenon.

⁹ We might use a weaker inequality: $m_\epsilon(S) - m_\epsilon(U) < \epsilon$. As far as I know this does not lead to any results.

A set of sufficient conditions is supplied by the theorem of Egoroff.¹⁰ We have already alluded to this theorem and the work of F. Riesz in deriving simple proofs. Perhaps the simplest of these proofs is that of the Monatshefte.¹¹ In the light of the first chapter it is in my opinion possible to get a better appreciation of this beautiful proof of Riesz.

As a first trivial remark¹² let us point out that *if all the sets $S - S(k, \nu)$ are closed, then Ψ contains S ; i. e., the sequence is uniformly convergent on the whole set S .*

Convergence on the part of the sequence $f_n(x)$ implies that $S - S(k, \nu)$ contains $S - S(k, \nu + 1)$, and $\prod_{\nu=1}^{\infty} \{S - S(k, \nu)\} = 0$ ¹³ for every k . Hence for any particular k it is true, since these sets are all closed, that there exists some ν_k such that $S - S(k, \nu_k) = 0$.

Pick $\nu_1, \nu_2, \nu_3, \dots$ as the sequence (ν) and consider the set $U(\nu) = \prod_{k=1}^{\infty} S(k, \nu_k)$.

Since $S - U(\nu) = \sum_{k=1}^{\infty} \{S - S(k, \nu_k)\} = 0$ it is clear that the sequence is uniformly convergent on the whole set S .

2. 2. One might make the intuitive remark that *if all the sets $S - S(k, \nu_k)$ are nearly closed then the sequence is uniformly convergent on very nearly the whole set S .*

In this connection let us make the following definition, keeping as close as possible to the familiar notions of classical analysis.

A set S is *approximately closed*¹⁴ if for every $\epsilon > 0$ there exists a set s such that $m_{\epsilon}(s) < \epsilon$ and $S - s$ is closed.

Before proceeding to discuss the above intuitive remark we shall state a few properties of approximately closed sets. All sets mentioned are assumed to be bounded.

2. 3. *The sum of two approximately closed sets is approximately closed.*

2. 4. *A bounded open set is approximately closed.*

2. 5. *The complement of an approximately closed set in an interval is approximately closed.*

¹⁰ D. Th. Egoroff, footnote ¹, p. 549, *loc. cit.*, p. 244.

¹¹ F. Riesz, footnote ² (iii), p. 549, *loc. cit.*, pp. 244-246.

¹² This same trivial remark forms the nucleus of the latest proof by Riesz, footnote ² (iii), p. 549, *loc. cit.*, p. 244.

¹³ $E = 0$ means that E is the null set.

¹⁴ An approximately closed set is clearly measurable, and conversely.

2.6. *The difference of two approximately closed sets is approximately closed.*

2.7. *The product of any number of approximately closed sets is approximately closed.*

2.8. *If the elements in a sequence of approximately closed sets have the character that each contains the next and their product is empty, then their exterior measures approach the limit zero.*

These properties are capable of immediate proof.¹⁵ By way of illustration we shall demonstrate the last.

Suppose that the sets are S_1, S_2, S_3, \dots and $\prod_1^\infty S_n = 0$. Given any $\epsilon > 0$ pick a series of positive terms $\epsilon_1 + \epsilon_2 + \epsilon_3 + \dots = \epsilon$. Now $S_n = C_n + s_n$ where C_n is closed and $m_e(s_n) < \epsilon_n$, for every positive integral value of n . Define $C_n^* = C_1 \cdot C_2 \cdot C_3 \cdot \dots \cdot C_n$. Then $\prod_1^\infty C_n^* = \prod_1^\infty C_n = 0$, and S_n contains C_n^* .

Since C_n^* contains C_{n+1}^* , and all of these sets are closed, after a certain n_0 all the sets C_n^* are empty. Hence for every $n > n_0$ we have $m_e(C_n^*) = 0$.

If x is not in C_n^* but is in $S_n = S_1 \cdot S_2 \cdot S_3 \cdot \dots \cdot S_n$, then x is not in some C_m , ($1 \leq m \leq n$) but is in all the S_k , ($k = 1, 2, 3, \dots, n$). Hence x is in s_m .

Therefore $C_n^* + (s_1 + s_2 + s_3 + \dots + s_n)$ contains S_n which contains C_n^* and so $S_n = C_n^* + s_n^*$ where $m_e(s_n^*) < \epsilon_1 + \epsilon_2 + \epsilon_3 + \dots + \epsilon_n$.

For $n > n_0$: $m_e(S_n) \leq m_e(C_n^*) + m_e(s_n^*) < 0 + \epsilon_1 + \epsilon_2 + \epsilon_3 + \dots + \epsilon_n < \epsilon$.

Hence $m_e(S_n) \rightarrow 0$.

2.9. We may now turn to the intuitive remark of 2.2 and show that if all the sets $S - S(k, \nu)$ are approximately closed then the sequence is approximately uniformly convergent.¹⁶

Proof. Given any $\epsilon > 0$ take a convergent series of positive terms $\epsilon_1 + \epsilon_2 + \epsilon_3 + \dots = \epsilon$.

Since $S - S(k, \nu)$ contains $S - S(k, \nu + 1)$ and $\prod_{\nu=1}^\infty \{S - S(k, \nu)\} = 0$ it follows from 2.8 that $m_e\{S - S(k, \nu)\} \rightarrow 0$ as $\nu \rightarrow \infty$, for every k .

¹⁵ The proofs are materially aided by two criteria: (i) If for every $\epsilon > 0$ there is an s of exterior measure $< \epsilon$ making $S + s$ approximately closed, then S is approximately closed; (ii) If for every ϵ , $S = S_\epsilon + s_\epsilon$ where S_ϵ is approximately closed and $m_e(s_\epsilon) < \epsilon$, then S is approximately closed.

¹⁶ F. Riesz, footnote ² (iii), p. 549, *loc. cit.*, p. 244.

Thus for every k there exists a ν_k such that

$$m_e\{S - S(k, \nu)\} < \epsilon_k.$$

Define $(\nu) : \nu_1, \nu_2, \nu_3, \dots$.

$$\begin{aligned} \text{Now } m_e\{S - U(\nu)\} &= m_e\{S - \prod_{k=1}^{\infty} S(k, \nu_k)\} = m_e\left\{\sum_{k=1}^{\infty} (S - S(k, \nu_k))\right\} \\ &\leq \sum_{k=1}^{\infty} m_e\{S - S(k, \nu_k)\} \leq \epsilon_1 + \epsilon_2 + \epsilon_3 + \dots = \epsilon. \end{aligned}$$

2.10. The problem of insuring approximate uniform convergence on the part of the sequence $f_n(x)$ thus resolves itself into the problem of insuring the sets $S - S(k, \nu)$ to be approximately closed. If we assume that the set S is approximately closed then by 2.6 the problem is further reduced to insuring that the sets $S(k, \nu)$ are approximately closed. This clearly follows if the functions are all continuous, but the condition is much too restrictive. As in the case of sets, we remain as close as possible to the familiar notions of classical analysis and define a class of functions which, while much more general, are close enough to continuous functions to carry us through.

A function $f(x)$ is said to be *approximately continuous*¹⁷ on a set S if for every $\epsilon > 0$ there exists a set s such that $m_e(s) < \epsilon$ and $f(x)$ is continuous on $S - s$ with respect to $S - s$.

We list some properties of approximately continuous functions which lend themselves to immediate proof.

2.11. *The sum, difference, product and quotient (if the denominator is different from zero except on a subset of measure 0) of two functions which are approximately continuous on a set S are again approximately continuous on S .*

2.12. *The absolute value of a function which is approximately continuous on S is approximately continuous on S .*

2.13. *If the function $f(x)$ is approximately continuous on an approximately closed set S and if c is any constant, then the subset of S on which the inequality $f(x) \leq c$ holds, is approximately closed.*

2.14. We may now show that if S is approximately closed and the functions $f_n(x)$ are approximately continuous, then the sets $S - S(k, \nu)$ are approximately closed.

¹⁷ The identity of functions approximately continuous on an interval and functions measurable on the same interval may be shown by using the theorem of Egoroff, or results of Borel and Hahn, footnote ² (iii), p. 549, *loc. cit.*, pp. 246-247.

Since S is approximately closed it follows from 2.6 that it is sufficient to show that the sets $S(k, \nu)$ are approximately closed.

If $s_{k,n,m}$ is the set of points of S where $|f_n(x) - f_m(x)| \leq 1/k$, then $S(k, \nu) = \prod_{n,m=\nu}^{\infty} s_{k,n,m}$. Hence by 2.7 it remains but to show that the sets $s_{k,n,m}$ are approximately closed. By 2.11 and 2.12 $|f_n(x) - f_m(x)|$ is approximately continuous and so by 2.13 $s_{k,n,m}$ is approximately closed.

2.15. Combining 2.9 and 2.14 we now have the following theorem:

If on an approximately closed set S we have a convergent sequence of approximately continuous functions $f_n(x)$, then the sequence is approximately uniformly convergent.

The theorem is of course valid if the convergence of the hypothesis is convergence almost everywhere. This is the celebrated theorem of Egoroff.

2.16. An immediate application is the following: *If $f_n(x)$ is a sequence of functions defined and approximately continuous on an approximately closed set S , and if the sequence converges almost everywhere to a function $f(x)$, then $f(x)$ is approximately continuous.*

The proof is immediate by reducing the considerations to a slightly smaller closed set on which all the functions are continuous and the convergence is uniform.

Hence, while in classical analysis it is *not* true that a sequence of *continuous functions* defined and convergent on a *closed set* is *uniformly convergent* and the limit function is *continuous*, the above statement becomes true if the word *approximately* is inserted before the words *continuous*, *closed* and *uniformly convergent*.

2.17. It is to be noticed that as a consequence of 2.5 a function which is *approximately continuous on an approximately closed set* may be made *approximately continuous on an interval*.

Suppose that $f(x)$ is approximately continuous on an approximately closed set S lying in the interval (a, b) . Then the function

$$f^*(x) = \begin{cases} f(x) & \text{on } S \\ 0 & \text{elsewhere in } (a, b) \end{cases}$$

is approximately continuous on (a, b) .

It follows that we need consider only functions which are approximately continuous on an interval. We now have the well known theorem:

A necessary and sufficient condition that a function $f(x)$ be approximately continuous on an interval (a, b) , is that there exist a sequence of continuous functions defined on (a, b) and converging almost everywhere in this interval to $f(x)$.

The sufficiency follows from 2.16.

To prove the necessity take any convergent series of positive terms $\eta_1 + \eta_2 + \eta_3 + \dots$. For every n it is true that there is an open set o_n such that $f(x)$ is continuous on $(a, b) - o_n = C_n$ and $m_e(o_n) < \eta_n$. Define a

function $c_n(x) = \begin{cases} f(x) & \text{on } C_n \\ \text{linear in the intervals of } o_n \text{ and taking on continuously} \\ & \text{the values } f(x) \text{ at the ends of these intervals.} \end{cases}$

Now $c_n(x)$ is continuous. Moreover, if ξ is in only a finite number of the sets o_n then $c_n(\xi) \rightarrow f(\xi)$, since then there is an $n(\xi)$ such that for all $n > n(\xi)$, ξ will be in C_n and so $c_n(\xi) = f(\xi)$.

The set of points in only a finite number of the sets is

$$Z = \left(\sum_{\nu=1}^{\infty} o_{\nu} \right) \cdot \left(\sum_{\nu=2}^{\infty} o_{\nu} \right) \cdot \left(\sum_{\nu=3}^{\infty} o_{\nu} \right) \cdot \dots$$

But $\sum_{\nu=n}^{\infty} o_{\nu}$ contains $\sum_{\nu=n+1}^{\infty} o_{\nu}$ and $m_e\left(\sum_{\nu=n}^{\infty} o_{\nu}\right) < \eta_n + \eta_{n+1} + \eta_{n+2} + \dots$. Since

the series is convergent $m_e(Z) = 0$ and $c_n(x) \rightarrow f(x)$ for almost every x .

This characterization of approximately continuous functions is of immediate importance, since now the theory of Lebesgue integration may be developed along lines similar to the work of Riesz,¹⁸ beginning with the well known concept of the integral of a continuous function. *The salient fact is that the Lebesgue theory of integration may be successfully presented with notions very similar to the familiar concepts of classical analysis (in fact within epsilon of these).*

2.18. Let us come back from the applications of the theorem of Egoroff to the first problem of this chapter. *The problem was to find conditions powerful enough to insure approximate uniform convergence.* One set of conditions has been given by Egoroff, and restricts not only the type of

¹⁸ F. Riesz, footnote ² (i), p. 549, *loc. cit.*

functions which make up the sequence $f_n(x)$, but also the set S on which they are defined. We find that *it is unnecessary to restrict the set S* . The proof of this fact is based on two auxiliary theorems, the first of which is a lemma of Sierpinski and Zygmund.¹⁹

AUXILIARY THEOREM I. *If a function $f(x)$ is continuous on a set S then there is an approximately closed set M containing S and a function $f^*(x)$ continuous on M such that $f^*(x) = f(x)$ on S .*

Proof. Consider the closure of S which is $S \cup S' = S^*$. Let M be the set of points of S^* where the saltus²⁰ of $f(x)$ is zero. It is clear that M contains S . Moreover, S^* is closed and so the subset S^{*}_k of its points where the saltus of $f(x)$ is $\geq 1/k$ is closed. But $M = \bigcap_{k=1}^{\infty} (S^* - S^{*}_k)$ and so it is approximately closed by 2.6 and 2.7. The limit of $f(x)$ at every point of M is unique. Define $f^*(x)$ to be equal to $f(x)$ on S and the limit of $f(x)$ at every point of $M - S$. M is the set and $f^*(x)$ the function required by the theorem.

AUXILIARY THEOREM II. *If a sequence of continuous functions $f_n(x)$ is defined on an approximately closed set M then the subset of M where the sequence converges is approximately closed.*

Proof. Define $\bar{f}_n(x) = \lim_{m \rightarrow \infty} \overline{f_n(x), f_{n+1}(x), f_{n+2}(x), \dots, f_m(x)}$ and $\underline{f}_n(x) = \lim_{m \rightarrow \infty} \underline{f_n(x), f_{n+1}(x), f_{n+2}(x), \dots, f_m(x)}$.²¹

These functions are approximately continuous by 2.16 for every value of n , and so $\limsup f_n(x) = \lim \bar{f}_n(x)$ and $\liminf f_n(x) = \lim \underline{f}_n(x)$ are also approximately continuous. Hence $(\limsup f_n(x) - \liminf f_n(x))$ is an approximately continuous function by 2.11, and it follows from 2.13 that the set of points where the above difference is zero is an approximately closed set. In other words the set of points of M where the limit exists is approximately closed.

We are now in a position to prove that *if a sequence of approximately continuous functions $f_n(x)$ is defined and convergent on any set S , then the sequence is approximately uniformly convergent*.

¹⁹ "Sur une fonction discontinue," *Fundamenta Mathematicae*, vol. 4 (1923), p. 317.

²⁰ The saltus of $f(x)$ at ξ is the least upper bound of the difference $[\limsup f(\dot{x}_n) - \liminf f(\dot{x}_n)]$ for all possible sequences $\{\dot{x}_n\}$ (chosen from points of S) \dot{x}_n and \ddot{x}_n converging to ξ .

²¹ See footnote 6, p. 551.

Proof. Given $\epsilon > 0$ take a sequence of positive terms $\epsilon_1 + \epsilon_2 + \epsilon_3 + \dots = \epsilon/2$. For every n there is an s_n of exterior measure $< \epsilon_n$ such that $f_n(x)$ is continuous on $S - s_n$. Hence the functions $f_n(x)$, ($n = 1, 2, 3, \dots$) are all continuous on the set $S - \Sigma s_n$ and $m_e(\Sigma s_n) < \epsilon/2$.

For every value of n , by Auxiliary Theorem I, there is an approximately closed set M_n containing $S - \Sigma s_n$ and a function $f_n^*(x)$ continuous on M_n such that $f_n^*(x) = f_n(x)$ on $S - \Sigma s_n$. Hence the functions $f_n^*(x)$, ($n = 1, 2, 3, \dots$) are all continuous on the approximately closed set $M = \Pi M_n$ and M contains $S - \Sigma s_n$.

By Auxiliary Theorem II the subset M^* of M on which the sequence $f_n^*(x)$ is convergent is approximately closed, and obviously contains $S - \Sigma s_n$.

By the theorem of Egoroff, given $\epsilon/2$ there is a set s of exterior measure $< \epsilon/2$ such that the convergence is uniform on $M^* - s$. Hence the convergence of the sequence $f_n^*(x)$ is uniform, *a fortiori*, on $S - (\Sigma s_n + s)$. Since $f_n^*(x) = f_n(x)$, ($n = 1, 2, 3, \dots$) on $S - \Sigma s_n$ and $m_e(\Sigma s_n + s) < \epsilon$ the sequence $f_n(x)$ is approximately uniformly convergent on the set S .

It appears, then, that *when we consider the elements of the class Ψ from the point of view of measure we come to the conclusion that as long as the functions of the sequence $f_n(x)$ are approximately continuous on S there is always a set U as close in measure to S as we please.*

THE OHIO STATE UNIVERSITY,
COLUMBUS, OHIO.

ON THE MOMENTUM PROBLEM FOR DISTRIBUTION FUNCTIONS IN MORE THAN ONE DIMENSION.

By E. K. HAVILAND.

The Hausdorff momentum problem¹ has recently been solved in the multi-dimensional case by Hildebrandt and Schoenberg.² In the present paper there will be treated the corresponding extension of the more general one-dimensional Hamburger¹ momentum problem; i. e., finding a necessary and sufficient condition for the existence of a distribution function³ ϕ such that

$$c_{nm} = \int \int_S x^n y^m d_{xy} \phi(E), \quad (n, m = 0, 1, 2, \dots),$$

where S denotes the entire (x, y) -plane and $\|c_{nm}\|$, $(n, m = 0, 1, 2, \dots)$, is a given real infinite matrix in which $c_{00} = 1$.

The majority of methods, in particular those based on Jacobi matrices and continued fractions, seem inapplicable in more than one dimension. However, the method of M. Riesz,⁴ developed from the ideas of F. Riesz in connection with linear functionals, can be extended to the multi-dimensional case, and the purpose of the present paper is to carry out that extension, the proofs being given, for convenience, in the case of two dimensions.

We consider the operation which makes correspond to any polynomial,

$$P(x, y) = \sum_{n=0}^N \sum_{m=0}^M a_{nm} x^n y^m,$$

the number

$$P_c = \sum_{n=0}^N \sum_{m=0}^M a_{nm} c_{nm},$$

where $\|c_{nm}\|$ is a given real infinite matrix. The operation is seen to be distributive. It is said to be non-negative if $P_c \geq 0$ provided $P(x, y)$ is non-negative, i. e., provided $P(x, y) \geq 0$ for all (x, y) , and in this case the matrix

¹ For references to literature on the momentum problem in one dimension, cf. M. H. Stone, *op. cit.*, pp. 613-614. References are collected at the end of the present paper.

² T. H. Hildebrandt and I. J. Schoenberg, *loc. cit.*

³ The monotone absolutely additive set function $\phi(E)$ is said to be a distribution function if $0 \leq \phi(E) \leq 1$ and $\phi(S) = 1$, where S denotes the whole (x, y) -plane. Cf. E. K. Haviland, *loc. cit.*, p. 627.

⁴ M. Riesz, *loc. cit.*, pp. 4-8.

$\|c_{nm}\|$ will be said to be non-negative. The result of this investigation is then given by the

THEOREM. *For the existence of a distribution function $\phi(E)$ such that*

$$(1) \quad \iint_S x^n y^m d_{xy} \phi(E) = c_{nm}, \quad (n, m = 0, 1, 2, \dots; c_{00} = 1),$$

it is necessary and sufficient that the matrix $\|c_{nm}\|$ be non-negative.

Proof. The necessary condition is immediately clear. For if a polynomial $P(x, y) \geq 0$ for every real (x, y) and if $\phi(E)$ is a distribution function, $\iint_S P(x, y) d_{xy} \phi(E) \geq 0$ and hence, if (1) is to hold, we must have $P_c \geq 0$.

To prove the sufficient condition, we let $P_{ij} : (\xi_i, \eta_j)$, $(i, j = 1, 2, \dots)$ be a denumerable set of points dense in S . In particular, we suppose them to be the intersections of sets of lines parallel to the coordinate axes and everywhere dense in the plane and let the functions $g_{ij}(x, y)$ be defined by $g_{ij}(x, y) = 1$ if $x < \xi_i$ and $y < \eta_j$; while $g_{ij}(x, y) = 0$ otherwise. The operation which makes P_c correspond to the polynomial $P(x, y)$ can be extended to the modul generated by finite linear combinations of $1, x, y, x^2, xy, y^2, \dots, g_{11}(x, y), g_{12}(x, y), g_{21}(x, y), \dots$ with real constants as coefficients in such a way that the operation remains distributive and non-negative, in the sense that to every non-negative function of this modul there corresponds a non-negative functional value.

This extension is made step by step. We consider first the modul A_1 generated by the various powers $x^m y^n$, $(m, n = 0, 1, 2, \dots)$, and by $g_{11}(x, y)$. To $g_{11}(x, y)$ we attach, as the value of the functional, a number γ_{11} , attaching at the same time to every finite linear combination of $1, x, y, \dots$, and $g_{11}(x, y)$ the corresponding combination of $c_{00} (= 1)$, c_{10} , c_{01} , \dots and γ_{11} . There is thus defined a distributive operation upon the modul A_1 .

In order that the operation be non-negative, γ_{11} will have to be so chosen that $\underline{\gamma}_{11} \leq \gamma_{11} \leq \bar{\gamma}_{11}$, where $\underline{\gamma}_{11}$ is the upper limit of the values which the operation makes correspond to all polynomials not greater than $g_{11}(x, y)$ for any (x, y) and $\bar{\gamma}_{11}$ is the lower limit of the values which the operation makes correspond to all polynomials not less than $g_{11}(x, y)$ for any (x, y) . That such polynomials actually exist may be seen from the fact that $f(x, y) \equiv 0$ belongs to the former class and $f(x, y) \equiv 1$ to the latter. The operation being distributive and non-negative in the modul of polynomials, we shall have⁵

⁵ The cases $\underline{\gamma}_{11} = \bar{\gamma}_{11}$ and $\underline{\gamma}_{11} < \bar{\gamma}_{11}$ are associated respectively with the determi-

$\gamma_{11} \leq \bar{\gamma}_{11}$. If $\gamma_{11} < \bar{\gamma}_{11}$, we choose, for definiteness, $\gamma_{11} = \bar{\gamma}_{11}$. In this way, the operation on the field A_1 is made non-negative as well as distributive, since then $P(x, y) + \mu g_{11}(x, y) \geq 0$ implies $P_c + \mu \gamma_{11} \geq 0$.

We next form the modul A_2 by adjoining to the generators of A_1 the function $g_{12}(x, y)$. To $g_{12}(x, y)$ we assign as its value the number γ_{12} , the lower limit of the values corresponding to those functions of the modul A_1 which are not less than $g_{12}(x, y)$ for any (x, y) . We thus obtain a distributive and non-negative operation defined over the modul A_2 . Continuing in this manner, we extend the operation to the moduls A_3, A_4, \dots obtained by the successive adjunction of the functions $g_{21}(x, y), g_{13}(x, y), \dots$ to the modul A_2 .

We then define a function $F(x, y)$ at the points $P_{ij} : (\xi_i, \eta_j)$ by the equation $F(\xi_i, \eta_j) = \gamma_{ij}$, where γ_{ij} denotes the value of the functional corresponding to the function $g_{ij}(x, y)$. This function $F(x, y)$ possesses, on the points P_{ij} , the monotone property in the sense of Radon.⁶ For suppose $\xi_{i_1} < \xi_{i_2}$ and $\eta_{j_1} < \eta_{j_2}$. Then from the definition of the functions $g_{ij}(x, y)$ it follows that for all (x, y)

$$g_{i_1 j_2}(x, y) - g_{i_1 j_1}(x, y) \leq g_{i_2 j_2}(x, y) - g_{i_2 j_1}(x, y).$$

Accordingly, as all these functions are included in some one of the moduls A_k (and, of course, in all succeeding moduls), it follows that

$$0 \leq \gamma_{i_2 j_2} - \gamma_{i_2 j_1} - \gamma_{i_1 j_2} + \gamma_{i_1 j_1},$$

i. e.,

$$(2) \quad 0 \leq F(\xi_{i_2}, \eta_{j_2}) - F(\xi_{i_2}, \eta_{j_1}) - F(\xi_{i_1}, \eta_{j_2}) + F(\xi_{i_1}, \eta_{j_1}).$$

Thus $F(x, y)$ possesses on the points P_{ij} the monotone property, q. e. d.

We shall next show that $F(-\infty, y) = F(x, -\infty) = 0$, where $F(-\infty, y) = \lim_{x \rightarrow -\infty} F(x, y)$, the points (x, y) belonging to the sequence $\{P_{ij}\}$ and the approach to the limit being uniform with respect to y , and a similar interpretation is to be placed on $F(x, -\infty)$. If $\xi_i < 0$, we have $g_{ij}(x, y) \leq \xi_i^{-2} x^2$ for all j and all (x, y) . It follows that

$$0 \leq F(\xi_i, \eta_j) = \gamma_{ij} \leq c_{20} \xi_i^{-2}$$

wherefore, as $\xi_i \rightarrow -\infty$, $\lim F(\xi_i, \eta_j) = 0$ uniformly for all η_j , q. e. d.

Similarly, if $\eta_j < 0$, we have $g_{ij}(x, y) \leq \eta_j^{-2} y^2$, so it may be shown in a

nateness or the non-determinateness of the momentum problem. For the one-dimensional case, cf. M. Riesz, *loc. cit.*, p. 9.

⁶ Cf. J. Radon, *loc. cit.* I, p. 1304.

similar manner that as $\eta_j \rightarrow -\infty$, $\lim F(\xi_i, \eta_j) = 0$ uniformly for all ξ_i . Again, $g_{ij}(x, y) \leq 1$, wherefore

$$0 \leq F(\xi_i, \eta_j) = \gamma_{ij} \leq c_{00} = 1.$$

With $F(x, y)$ there may be associated an interval function $\psi(I)$ defined for an everywhere dense⁷ set of intervals

$$I : (\xi_{i_1} \leq x < \xi_{i_2}; \eta_{j_1} \leq y < \eta_{j_2})$$

$$\text{by} \quad \psi(I) = F(\xi_{i_2}, \eta_{j_2}) - F(\xi_{i_2}, \eta_{j_1}) - F(\xi_{i_1}, \eta_{j_2}) + F(\xi_{i_1}, \eta_{j_1}),$$

and the definition continues to hold when $\xi_{i_1} = \eta_{j_1} = -\infty$, in which case $\psi(I_{i_2j_2}) = F(\xi_{i_2}, \eta_{j_2})$, where $I_{i_2j_2} : (-\infty < x < \xi_{i_2}; -\infty < y < \eta_{j_2})$. From its definition $\psi(I)$ is seen to be additive and we have already seen from equation (2) that it possesses the monotone property. In consequence, as $h \rightarrow 0$, $k \rightarrow 0$, $\lim F(\xi - h, \eta - k)$ exists, provided $h \geq 0$, $k \geq 0$ and the points $(\xi - h, \eta - k)$, with perhaps the exception of (ξ, η) , belong to the everywhere dense set of points for which F is defined.

We now extend the definition of F to all points (x, y) not belonging to the given everywhere dense set by setting $F(x, y) = \lim_{h=0, k=0} F(x-h, y-k)$, where $(x-h, y-k)$ is a point of the everywhere dense set and $h \geq 0$, $k \geq 0$, and we define $\psi(I)$ for any interval $I : (x_1 \leq x < x_2; y_1 \leq y < y_2)$ by

$$(3) \quad \psi(I) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1).$$

It is seen from the definition that $\psi(I)$ is additive. Moreover, it is monotone, for there exist points (ξ_{i_1}, η_{j_1}) , (ξ_{i_2}, η_{j_2}) , (ξ_{i_2}, η_{j_1}) , (ξ_{i_1}, η_{j_2}) such that for any $x_1 < x_2$ and $y_1 < y_2$

$$0 \leq F(x_m, y_n) - F(\xi_{i_m}, \eta_{j_n}) < \epsilon, \quad (m, n = 1, 2),$$

and this, together with (2), implies

$$0 \leq F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1).$$

Similarly, it may be shown that $F(x, y)$, as thus defined for all points of the plane, is such that $F(-\infty, y) = F(x, -\infty) = 0$, while $F(x, y) \leq 1$ for all (x, y) . In fact, it will appear later that $F(+\infty, +\infty) = 1$. Hence $\psi(I)$ is a bounded additive monotone non-decreasing interval function. It follows⁸

⁷ Cf. E. K. Haviland, *loc. cit.*, p. 628, Definition 4.

⁸ The proof is similar to that given by J. Radon, *loc. cit.* II, p. 1093.

that its discontinuities, if any, fall upon a denumerable set of lines parallel to the coördinate axes. In consequence, there is an everywhere dense set of points (ξ'_i, η'_j) , which may be taken to be the intersections of two everywhere dense sets of lines parallel to the coördinate axes, such that

$$\lim_{h=0, k=0} F(\xi'_i - h, \eta'_j - k) = F(\xi'_i, \eta'_j), \quad h \geq 0, k \geq 0,$$

where the points $(\xi'_i - h, \eta'_j - k)$ belong to the same everywhere dense set as does (ξ'_i, η'_j) . Then there exists⁹ a bounded monotone absolutely additive set function $\phi(E)$ whose corresponding point function, $G(x, y)$, coincides with $F(x, y)$ on the everywhere dense set of points (ξ'_i, η'_j) . Moreover, $G(x, y)$ and $F(x, y)$ have the same discontinuity points and are equal at all other points, i. e., $\psi(I)$ and $\phi(E)$ are equal on all their non-singular rectangles. We shall show that $\phi(E)$ is a solution of the momentum problem belonging to the preassigned matrix $\|c_{nm}\|$.

To this end, we consider a monomial¹⁰ x^ny^m , where n, m are arbitrary non-negative integers. Let $2r$ be a fixed even number greater than $n + m$, and choose $-T_1 < 0$ and $T_2 > 0$ so that they belong to the set ξ_1, ξ_2, \dots , and $-T'_1 < 0$ and $T'_2 > 0$ so that they belong to the set η_1, η_2, \dots . Furthermore, T_1, T_2, T'_1, T'_2 shall be so large that

$$|x^ny^m| < \epsilon(x^{2r} + y^{2r})$$

outside the rectangle $R : (-T_1 \leq x < T_2; -T'_1 \leq y < T'_2)$ and on its boundary, ϵ being a fixed arbitrarily small positive quantity. We divide R by lines $x_1 = \xi_{i_1} = -T_1, x_2 = \xi_{i_2}, \dots, x_{p+1} = \xi_{i_{p+1}} = T_2$ and $y_1 = \eta_{j_1} = -T'_1, y_2 = \eta_{j_2}, \dots, y_{q+1} = \eta_{j_{q+1}} = T'_2$ into a set of rectangles

$$R_{kl} : (x_k \leq x < x_{k+1}; y_l \leq y < y_{l+1})$$

in each of which the oscillation of x^ny^m is less than ϵ' , where ϵ' is another fixed arbitrarily small positive quantity. Let (X_k, Y_l) be a point in the interior of the rectangle R_{kl} . We then form the step function $v(x, y)$ which vanishes outside R and which takes the value $X_k^n Y_l^m$ in R_{kl} . Then for every (x, y)

$$v(x, y) - \epsilon' - \epsilon(x^{2r} + y^{2r}) < x^ny^m < v(x, y) + \epsilon' + \epsilon(x^{2r} + y^{2r}).$$

Since the function $v(x, y)$ belongs to one of the moduls A_1, A_2, \dots (and

⁹ Cf. E. K. Haviland, *loc. cit.*, p. 651, and the references there given.

¹⁰ The proof holds also for an arbitrary polynomial.

hence to any subsequent modul), the functional operation is defined for it, and if to $v(x, y)$ corresponds the functional value v_c ,

$$(4) \quad v_c - \epsilon' - \epsilon(c_{2r,0} + c_{0,2r}) \leq c_{nm} \leq v_c + \epsilon' + \epsilon(c_{2r,0} + c_{0,2r}).$$

Furthermore,

$$v(x, y) = \sum_{k=1}^p \sum_{l=1}^q X_k^n Y_l^m [g_{i_{k+1}j_{l+1}}(x, y) - g_{i_{k+1}j_l}(x, y) - g_{i_kj_{l+1}}(x, y) + g_{i_kj_l}(x, y)].$$

Hence, as $\gamma_{i_kj_l} = F(\xi_{i_k}, \eta_{j_l}) = F(x_k, y_l)$, we have by (3)

$$v_c = \sum_{k=1}^p \sum_{l=1}^q X_k^n Y_l^m \psi(R_{kl}).$$

The inequality (4) can then be written

$$(5) \quad \sum_{k=1}^p \sum_{l=1}^q X_k^n Y_l^m \psi(R_{kl}) - \epsilon' - \epsilon(c_{2r,0} + c_{0,2r}) \leq c_{nm} \\ \leq \sum_{k=1}^p \sum_{l=1}^q X_k^n Y_l^m \psi(R_{kl}) + \epsilon' + \epsilon(c_{2r,0} + c_{0,2r}).$$

Now $x^n y^m$ is continuous in x and y together in every rectangle and $\psi(I)$ is a bounded monotone additive interval function. Hence,¹¹ as the diameter of the R_{kl} approaches zero,

$$\lim_{k \rightarrow 1} \sum_{k=1}^p \sum_{l=1}^q X_k^n Y_l^m \psi(R_{kl}) = \iint_R x^n y^m d_{xy} \psi(I).$$

At the same time, $\epsilon' \rightarrow 0$, so that from (5) we obtain

$$\iint_R x^n y^m d_{xy} \psi(I) - \epsilon(c_{2r,0} + c_{0,2r}) \\ \leq c_{nm} \leq \iint_R x^n y^m d_{xy} \psi(I) + \epsilon(c_{2r,0} + c_{0,2r}).$$

Let $T_1, T'_1, T_2, T'_2 \rightarrow +\infty$ and $\epsilon \rightarrow 0$. Then

$$\iint_S x^n y^m d_{xy} \psi(I) = c_{nm}.$$

As, however, for any arbitrarily large non-singular rectangle R_1 of ϕ and ψ ,

$$\iint_{R_1} x^n y^m d_{xy} \psi(I) = \iint_{R_1} x^n y^m d_{xy} \phi(E),$$

it follows that¹²

¹¹ Cf., e. g., S. Bochner, *loc. cit.*, p. 391.

¹² This follows directly if m and n are both even. Otherwise one has first to use

$$\iint_S x^n y^m d_{xy}\phi(E) = c_{nm}, \quad \text{q. e. d.}$$

The distribution function ϕ whose existence is thus established is not necessarily uniquely determined. Sufficient conditions that ϕ be uniquely determined by its momenta c_{nm} have been found by V. Romanovsky¹³ and by the present author.¹⁴ In particular, ϕ is uniquely determined if the c_{nm} are such that

$$\left(\sum_{\nu=0}^{2n} \binom{2n}{\nu} |c_{2n-\nu, \nu}| \right)^{1/(2n)} = o(n),$$

and the present author has shown¹⁴ that this condition is almost necessary in that ϕ may not be uniquely determined if $o(n)$ is replaced by $o(n^{1+\epsilon})$.

REFERENCES.

-
- S. Bochner, "Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse," *Mathematische Annalen*, vol. 108 (1933), pp. 378-410.
 E. K. Haviland, "On the theory of absolutely additive distribution functions," *American Journal of Mathematics*, vol. 56 (1934), pp. 625-658.
 T. H. Hildebrandt and I. J. Schoenberg, "On linear functional operators and the moment problem for a finite interval in one or several dimensions," *Annals of Mathematics*, ser. 2, vol. 34 (1933), pp. 317-328.
 J. Radon, I: "Theorie und Anwendungen der absolut additiven Mengenfunktionen," *Sitzungsberichte der mathematischen-naturwissenschaftlichen Klasse der Kaiserl. Akademie zu Wien*, vol. 122 (1913), pp. 1295-1438; II: "Über lineare Funktionaltransformationen und Funktionalgleichungen," *ibid.*, vol. 128 (1919), pp. 1083-1121.
 M. Riesz, "Sur le problème des moments," *Arkiv för Matematik, Astronomi och Fysik*, vol. 17 (1923), No. 16.
 V. Romanovsky, "Sur un théorème limite du calcul des probabilités," *Recueil Mathématique de la Société Mathématique de Moscou*, vol. 36 (1929), pp. 36-64.
 M. H. Stone, "Linear transformations in Hilbert space," *American Mathematical Society Colloquium Publications*, vol. 15.

THE JOHNS HOPKINS UNIVERSITY.

the inequality of Schwarz. For the existence of the integral with respect to ϕ , cf. J. Radon, *loc. cit.* I, pp. 1322-1324. Notice that if $n = m = 0$, we obtain $\phi(S) = c_{00} = 1$.

¹³ V. Romanovsky, *loc. cit.*, p. 47, § 3.

¹⁴ E. K. Haviland, *loc. cit.*, p. 634. At the time of publishing that paper, the author was not aware that a proof of his Theorem I, under restricted conditions, and of the sufficient condition in his Theorem II had previously been given by Romanovsky. Romanovsky's statement of this sufficient condition differs somewhat from the present author's but the two proofs are effectively the same.

A NOTE ON A PROPERTY OF FOURIER-STIELTJES TRANSFORMS IN MORE THAN ONE DIMENSION.

By E. K. HAVILAND.

If $\rho(\xi)$ is monotone in $[-\infty, +\infty]$ and $\rho(-\infty) = 0$, $\rho(+\infty) = 1$; if $\Lambda(t; \rho) = \int_{-\infty}^{+\infty} \exp(it\xi) d\rho(\xi)$, where $-\infty < t < +\infty$, denotes the Fourier-Stieltjes transform of ρ ; and if, finally, $\mathfrak{M}(f(\cdot)) = \lim_{T \rightarrow +\infty} (2T)^{-1} \int_{-T}^T f(t) dt$, it is known¹ that $\mathfrak{M}(|\Lambda(\cdot; \rho)|^2) = \sum |\Delta_k|^2$ where $\Delta_k = \rho(\xi_k + 0) - \rho(\xi_k - 0)$ and the summation is taken over all the (at most denumerable) discontinuities of ρ . In particular, if ρ is continuous,

$$(1) \qquad \mathfrak{M}(|\Lambda(\cdot; \rho)|^2) = 0.$$

In the case of more than one dimension, the discontinuity points need no longer be denumerable, so the question arises as to what then corresponds to the foregoing result. It turns out that in the multi-dimensional case a similar result holds, the point spectrum, as defined in an earlier paper,² playing the rôle of the discontinuities in the one-dimensional case, while the "mild" discontinuity points, i. e., those not occurring in the point spectrum, play no rôle at all. More precisely, it will be shown that in more than one dimension,³ if $\phi(E)$ be a distribution function,⁴

¹ This result was stated, without proof, by Paul Lévy, *Calcul de probabilités* (Paris, 1925), p. 171. For a proof, cf. I. Schoenberg, "Über total monotone Folgen mit stetiger Belungungsfunktion," *Mathematische Zeitschrift*, vol. 30 (1929), pp. 761-767, where reference is made to a paper of N. Wiener. Since then, (1) has often been rediscovered in connection with the unitary dynamics of Carleman and Koopman and with the statistical considerations of Khintchine. Cf. also A. Wintner and E. K. Haviland, "On the Fourier-Stieltjes transform," *American Journal of Mathematics*, vol. 56 (1934), pp. 4-5.

² Cf. E. K. Haviland, "On the theory of absolutely additive distribution functions," *American Journal of Mathematics*, vol. 56 (1934), p. 654.

³ For convenience, we give the proof in the two-dimensional case.

⁴ The monotone absolutely additive set function $\phi(E)$ is said to be a distribution function if $0 \leq \phi(E) \leq 1$ and $\phi(S) = 1$, where S denotes the whole plane. Cf. E. K. Haviland, *ibid.*, p. 627. For the definition of integrals with respect to such functions, cf. J. Radon, "Theorie und Anwendungen der absolut additiven Mengenfunktionen," *Sitzungsberichte der mathematischen-naturwissenschaftlichen Klasse der Kaiserl. Akademie zu Wien*, vol. 122 (1913), pp. 1322-1324.

$$(2) \quad \mathfrak{M}(|\Lambda(s, t; \phi)|^2) = \sum_{k=1}^{\infty} [\phi(P_k)]^2,$$

where

$$(3) \quad \Lambda(s, t; \phi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp\{i(sx + ty)\} d_{xy} \phi(E)$$

and

$$\mathfrak{M}(f(s, t)) = \lim_{T=\infty, U=\infty} (2T)^{-1} (2U)^{-1} \int_{-T}^T \int_{-U}^U f(s, t) ds dt,$$

and the summation on the right of (2) is taken over all points P_k of the point spectrum of ϕ .

While the result (2) is analogous to that in the one-dimensional case, it is not obvious from the latter, since in two or more dimensions the singularities of a monotone function are essentially more complicated than those of such a function in a single dimension, where *all* discontinuity points belong to the point spectrum.

The proof of (2) is as follows: If $\phi(E)$ be a distribution function, we define a set function $\bar{\phi}(E)$ by setting

$$(4) \quad \bar{\phi}(E) = \phi(-E),$$

where $-E$ is the set symmetric to E with respect to the origin. Then $\bar{\phi}(E)$ is a distribution function, and,⁵ by virtue of the Convolution Theorem for Fourier-Stieltjes transforms,⁶ $\Lambda(s, t; \phi * \bar{\phi}) = \Lambda(s, t; \phi) \cdot \Lambda(s, t; \bar{\phi})$; or, since $\Lambda(s, t; \phi)$ and $\Lambda(s, t; \bar{\phi})$ are conjugated complex quantities in virtue of (3) and (4),

$$|\Lambda(s, t; \phi)|^2 = \Lambda(s, t; \phi * \bar{\phi}).$$

Consequently

$$(5) \quad \mathfrak{M}(|\Lambda(s, t; \phi)|^2) = \mathfrak{M}(\Lambda(s, t; \phi * \bar{\phi}))$$

and we need examine only the latter.

It is now to be shown that if $\psi(E)$ be a distribution function whose point spectrum is vacuous, then $\mathfrak{M}(\Lambda(s, t; \psi)) = 0$. Since the contribution of the integration domain $S - R$ to $\Lambda(s, t; \psi)$, where S represents the entire (x, y) -plane and R an arbitrary rectangle in that plane having its sides parallel to the coördinate axes, is in absolute value less than ϵ for all (s, t) provided R is sufficiently large, it is sufficient to prove that for any fixed R and for sufficiently large values of T, U

⁵ $\psi_1 * \psi_2$ denotes the symbolical product (Faltung or convolution) of ψ_1 and ψ_2 . Cf. E. K. Haviland, *loc. cit.*, p. 651, Theorem IV.

⁶ Cf. E. K. Haviland, *ibid.*, Theorem V.

$$\left| (4TU)^{-1} \int_{-T}^T \int_{-U}^U \left\{ \iint_R \exp[i(sx + ty)] d_{xy}\psi(E) \right\} ds dt \right| < \epsilon.$$

We begin the proof of this statement by observing that the expression beneath the absolute value signs may be written as

$$J = (4TU)^{-1} \int_{-T}^T \int_{-U}^U \left\{ \int_{-M}^M \int_{-N}^N \exp[i(sx + ty)] d_{xy}\psi(E) \right\} ds dt,$$

and it is permissible to invert the order of integration,⁷ obtaining

$$\begin{aligned} J &= \int_{-M}^M \int_{-N}^N \left\{ (2T)^{-1} \int_{-T}^T e^{isx} ds \cdot (2U)^{-1} \int_{-U}^U e^{ity} dt \right\} d_{xy}\psi(E) \\ &= \int_{-M}^M \int_{-N}^N \frac{\sin Tx}{Tx} \frac{\sin Uy}{Uy} d_{xy}\psi(E) \\ &= \int \int_{\text{I}} + \int \int_{\text{II}} + \int \int_{\text{III}} + \int \int_{\text{IV}} + \int \int_{\text{V}}, \end{aligned}$$

where

- I: $(-M \leq x < -\delta; -N \leq y \leq N)$, II: $(-\delta \leq x \leq \delta; \delta < y \leq N)$,
 III: $(\delta < x \leq M; -N \leq y \leq N)$, IV: $(-\delta \leq x \leq \delta; -N \leq y < -\delta)$,
 V: $(-\delta \leq x \leq \delta; -\delta \leq y \leq \delta)$.

Now $\left| \int \int_{\text{V}} \right| \leq \int \int_{\text{V}} d_{xy}\psi(E)$ and if $(0, 0)$ is not a point of the point spectrum of ψ , this last expression may be made less than $\epsilon/5$ in absolute value by taking δ sufficiently small. δ being thus fixed, it is easily seen that

$$\left| \int \int_{\text{I}} \right| \leq (T\delta)^{-1} \int \int_{\text{I}} d_{xy}\psi(E) < \epsilon/5$$

and

$$\left| \int \int_{\text{III}} \right| \leq (T\delta)^{-1} \int \int_{\text{III}} d_{xy}\psi(E) < \epsilon/5$$

provided T is sufficiently large. Similarly,

$$\left| \int \int_{\text{II}} \right| \leq (U\delta)^{-1} \int \int_{\text{II}} d_{xy}\psi(E) < \epsilon/5$$

and

$$\left| \int \int_{\text{IV}} \right| \leq (U\delta)^{-1} \int \int_{\text{IV}} d_{xy}\psi(E) < \epsilon/5$$

if U is sufficiently large. Consequently, if the point spectrum of ψ is vacuous, $\Re(\Lambda(s, t; \psi)) = 0$.

⁷ Cf. E. K. Haviland, *ibid.*, p. 640.

Furthermore,⁸ every absolutely additive set function ϕ of bounded total variation is the sum of two functions, say ϕ_1 and ϕ_2 , of which the former has a vacuous point spectrum, while the latter is purely discontinuous (i. e., its spectrum coincides with its point spectrum). Then from the definition of ϕ it follows that $\bar{\phi} = \bar{\phi}_1 + \bar{\phi}_2$ and

$$\begin{aligned}\mathfrak{M}(\Lambda(s, t; \phi * \bar{\phi})) &= \mathfrak{M}(\Lambda(s, t; (\phi_1 + \phi_2) * (\bar{\phi}_1 + \bar{\phi}_2))) \\ &= \mathfrak{M}(\Lambda(s, t; \phi_1 * \bar{\phi}_1)) + \mathfrak{M}(\Lambda(s, t; \phi_1 * \bar{\phi}_2)) \\ &\quad + \mathfrak{M}(\Lambda(s, t; \bar{\phi}_1 * \phi_2)) + \mathfrak{M}(\Lambda(s, t; \phi_2 * \bar{\phi}_2)).\end{aligned}$$

But the point spectra of $\phi_1 * \bar{\phi}_1$, $\phi_1 * \bar{\phi}_2$ and $\bar{\phi}_1 * \phi_2$ are vacuous by the addition rule⁹ for point spectra, so that the first three terms in the last member of the preceding equation vanish, and

$$\begin{aligned}(6) \quad \mathfrak{M}(\Lambda(s, t; \phi * \bar{\phi})) &= \mathfrak{M}(\Lambda(s, t; \phi_2 * \bar{\phi}_2)) \\ &= \mathfrak{M}(\Lambda(s, t; \phi_2) \cdot \Lambda(s, t; \bar{\phi}_2)) = \mathfrak{M}(\Lambda(s, t; \phi_2) \cdot \Lambda(-s, -t; \phi_2)).\end{aligned}$$

Let points of the point spectrum of ϕ , i. e., of ϕ_2 , be $P_k : (x_k, y_k)$. As they are at most denumerable, it follows that

$$\Lambda(s, t; \phi_2) = \sum_{k=1}^{\infty} \exp[i(sx_k + ty_k)] \phi(P_k)$$

and

$$\Lambda(-s, -t; \phi_2) = \sum_{k=1}^{\infty} \exp[-i(sx_k + ty_k)] \phi(P_k).$$

Substituting in (6) and taking account of (5), we find

$$\mathfrak{M}(|\Lambda(s, t; \phi)|^2) = \mathfrak{M}\left(\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \exp[i\{s(x_j - x_k) + t(y_j - y_k)\}]\phi(P_j)\phi(P_k)\right).$$

On taking the mean value, all terms for which $j \neq k$ disappear, so that the right-hand side of the preceding equation becomes $\sum_{k=1}^{\infty} [\phi(P_k)]^2$ which proves (2).

THE JOHNS HOPKINS UNIVERSITY.

⁸ Cf. H. Hahn, *Theorie der reellen Funktionen* (Berlin, 1921), p. 414, Theorem XV.

⁹ Cf. E. K. Haviland, *loc. cit.*, p. 654, Theorem VI.

THE THEORY OF THE SECOND VARIATION FOR THE NON-PARAMETRIC PROBLEM OF BOLZA.¹

By WILLIAM T. REID.

1. **Introduction.** The non-parametric problem of Bolza in the calculus of variations is that of finding in a class of arcs

$$(1.1) \quad y_i = y_i(x) \quad (i = 1, \dots, n; x_1 \leq x \leq x_2),$$

satisfying the differential equations and end conditions

$$(1.2) \quad \phi_\alpha[x, y, y'] = 0 \quad (\alpha = 1, \dots, m < n),$$

$$(1.3) \quad \psi_\gamma[x_1, y(x_1), x_2, y(x_2)] = 0 \quad (\gamma = 1, \dots, p \leq 2n + 2),$$

one which minimizes an expression of the form

$$(1.4) \quad I = G[x_1, y(x_1), x_2, y(x_2)] + \int_{x_1}^{x_2} f(x, y, y') dx.$$

Sufficient conditions for the problem of Bolza have been given by Morse [III]² and Bliss [IV] for extremal arcs that are not only normal relative to the end conditions but also normal on sub-intervals. Recently, sufficient conditions have been obtained under weaker assumptions by Hestenes [X], who has replaced the usual condition of Mayer by a new condition in terms of a certain quadratic form involving the solutions of the accessory equations. Hestenes has not only been able to discard the hypothesis of normality on sub-intervals, but has also obtained sufficient conditions for an extremal arc with multipliers of the form $\lambda_0 = 1$, $\lambda_\alpha(x)$ which is not necessarily normal relative to the end conditions.

The principal result of the present paper is the following theorem:

THEOREM A. *If $E_{12} : y_i = y_i(x)$, $\lambda_0 = \text{constant}$, $\lambda_\alpha(x)$, $x_1 \leq x \leq x_2$, is an extremal arc which satisfies the strengthened Glebsch condition, and on which there is no point conjugate to the point 1, then there exists a family*

¹ Presented to the American Mathematical Society, September 5, 1934.

² Roman numerals in brackets refer to the bibliography at the end of this paper. Only papers to which direct reference is made in the present paper are listed. For a more extensive bibliography the reader is referred to that given by Hestenes at the end of [X].

of n mutually conjugate accessory extremals $\eta_{ij}(x)$, $\zeta_{ij}(x)$ ($j = 1, \dots, n$) such that $|\eta_{ij}(x)| \neq 0$ on x_1x_2 .

This theorem is fundamental in the construction of a field of extremal imbedding a given extremal, and has been proved by several authors under additional assumptions of normality.³ Theorem A has been proved by Morse by an extension of the methods which he used in [II].⁴ *The chief significance of the independent proof here given is that it is a direct generalization of the method used by Bliss when the extremal arc satisfies additional normality conditions* [I, pp. 729, 736], *and hence is more intimately related to the methods usually used in the simpler problems of the calculus of variation than the methods of Morse and Hestenes.*

Certain general properties of accessory extremals are discussed in § 2 of this paper, and Theorem A is established in § 3. In § 4 there are proved by the use of the results of § 3, further results concerning the existence of families of accessory extremals satisfying the condition of Theorem A. In particular, Theorem 4.2 gives a rather elegant method for determining such a family of accessory extremals. It is to be noted, however, that the proof of the interesting result of Theorem 4.3 is independent of the results of § 3. Finally, in § 5 there is discussed briefly the relation of Theorem A to sufficiency theorems for the problems of Bolza and Mayer.

Throughout the paper, the coefficients of (1.2), (1.3), and (1.4) are supposed to satisfy the hypotheses usually made {see [III], [IV] and [X]}

2. Accessory extremals. For an extremal arc $E_{12}: y_i = y_i(x)$, $\lambda_0 = \text{constant}$, $\lambda_a(x)$, $x_1 \leq x \leq x_2$, let

$$\begin{aligned} F &= \lambda_0 f(x, y, y') + \lambda_a(x) \phi_a(x, y, y'), \\ (2.1) \quad 2\omega[x, \eta, \eta'] &= F_{y', y', \eta'} \eta_j' \eta_j' + 2F_{y', y, \eta'} \eta_j' + F_{y, y, \eta'} \eta_j, \\ \Phi_a[x, \eta, \eta'] &= \phi_{ay', \eta'} \eta_j' + \phi_{ay, \eta_j} \end{aligned} \quad (\alpha = 1, \dots, m)$$

The coefficients of ω and ϕ_a are supposed to have as arguments the function $y_i(x)$, λ_0 , $\lambda_a(x)$ belonging to E_{12} . It will also be supposed that E_{12} satisfies the strengthened Clebsch condition [IV, p. 264]. As usual, this condition will be denoted as III'. If we set

³ See [I], pp. 729, 736; [II]; [VII, p. 320; and [X], pp. 804, 807.

⁴ I was not aware that Morse had proved this result until the date upon which I presented my proof to the American Mathematical Society. Morse's paper has since appeared in the *Transactions of the American Mathematical Society*, vol. 37 (1935) pp. 147-160. Hestenes has informed me that subsequent to my proof of Theorem A he also proved this result by the use of the formulation of the Mayer condition which he has used in [X].

$$(2.2) \quad \Omega[x, \eta, \eta', \mu] = \omega[x, \eta, \eta'] + \mu_\alpha \Phi_\alpha[x, \eta, \eta'],$$

then the system of accessory differential equations is

$$(2.3) \quad (d/dx)\Omega_{\eta'_i} - \Omega_{\eta_i} = 0, \quad \Phi_\alpha = 0 \quad (i = 1, \dots, n; \alpha = 1, \dots, m).$$

By the introduction of the canonical variables $\xi_i = \Omega_{\eta'_i}[x, \eta, \eta', \mu]$, this system is seen to be equivalent to a system of $2n$ linear differential equations of the first order of the form

$$(2.3') \quad \eta'_i = A_{ij}(x)\eta_j + B_{ij}(x)\xi_j, \quad \xi'_i = C_{ij}(x)\eta_j - A_{ji}(x)\xi_j.$$

The coefficients in (2.3') are continuous on x_1x_2 , $\|B_{ij}\|$ and $\|C_{ij}\|$ are symmetric matrices, and $\|B_{ij}\|$ is of rank $n - m$ on x_1x_2 .⁵ We shall say that a set of functions $\eta_i(x)$, $\xi_i(x)$ which are of class O' , and which satisfy (2.3') on x_1x_2 is an *accessory extremal*.⁶

If the extremal E_{12} is a minimizing arc which is normal on every sub-interval x_3x_2 of x_1x_2 it has been shown by Bliss that there can be no point conjugate to 1 on E_{12} between 1 and 2.⁷ We shall use IV_0 to denote this necessary condition, and IV'_0 to denote the condition that there is no value x_3 such that $x_1 < x_3 \leq x_2$ and defining a point 3 conjugate to 1.

We shall say that the order of anormality⁸ of E_{12} on a sub-interval t_1t_2 of x_1x_2 is equal to r if on this sub-interval there are exactly r linearly independent accessory extremals $\eta_i = u_{ik}(x)$, $\xi_i = v_{ik}(x)$ ($k = 1, \dots, r$) with $u_{ik}(x) \equiv 0$ on t_1t_2 .

The following properties of accessory extremals will be given without proof.

PROPERTY 1°. *The order of anormality of E_{12} on a given sub-interval is at most m .*

⁵ See, for example, [VIII], §§ 3 and 4.

⁶ The terminology *accessory differential system* for the system (2.3) is due to von Escherich. The problem of minimizing the second variation in a class of arcs satisfying the equations of variation has been called the *accessory minimum problem* [III, and X], and the associated boundary value problem has been termed the *accessory boundary value problem* [III, VIII and X]. On the other hand, a set of functions $\eta_i(x)$ belonging to a solution η_i, μ_α of (2.3), or to a solution η_i, ξ_i of (2.3'), has been called a *secondary extremal*. [III, and X]. It seems more consistent to either speak of an *accessory extremal*, or else to use the terms *secondary differential system*, *secondary minimum problem*, *secondary boundary value problem*, and *secondary extremal*. Due to the priority of the term *accessory* for the differential system, the present author has adopted the phrase *accessory extremal* in the sense defined above.

⁷ See [I], p. 725. The reader is referred to [I] for the definition of conjugate point.

⁸ This terminology has been used by Hestenes, [X], p. 799.

PROPERTY 2°. If the order of anormality of E_{12} on a sub-interval $t_1 t_2$ is r , and $\eta_i \equiv 0$, $\xi_i = v_{ik}(x)$ ($k = 1, \dots, r$) are linearly independent accessory extremals on this sub-interval, then for arbitrary admissible variations $\eta_i(x)$ and arbitrary points x', x'' of $t_1 t_2$, we have

$$v_{ik}(x)\eta_i(x) \Big|_{x=x'}^{x=x''} = 0 \quad (k = 1, \dots, r).$$

We shall denote by $r(x)$ the order of anormality of E_{12} on the sub-interval $x_1 x$ ($x_1 < x \leq x_2$) of $x_1 x_2$. The function $r(x)$ is seen to be monotone non-increasing on $x_1 < x \leq x_2$. In view of Property 1°, we have

PROPERTY 3°. There exists a constant d such that $0 < d \leq x_2 - x_1$, and $r(x)$ is constant on $x_1 < x \leq x_1 + d$.

The following property is a consequence of the continuity of the solutions η_i, ξ_i of (2.3'):

PROPERTY 4°. If $x_1 < x_3 \leq x_2$, there exists a δ such that $0 < \delta < x_3 - x_1$ and $r(x) = r(x_3)$ on $x_3 - \delta \leq x \leq x_3$.

In view of the above properties, it is seen that $r(x)$ has at most m points of discontinuity on $x_1 < x \leq x_2$. We shall denote these points by t_1, \dots, t_g , where $x_1 < t_g < t_{g-1} < \dots < t_1 < x_2$. For convenience, let $t_{g+1} = x_1$, $t_0 = x_2$, and $r_q = r(t_q)$ ($q = 0, 1, \dots, g$).

It is readily seen that one may choose a family of accessory extremals $u_{ij}(x), v_{ij}(x)$ ($j = 1, \dots, n$) such that

$$(2.4) \quad \begin{aligned} u_{ij}(x_1) &= 0, \quad u_{i\theta}(x) \equiv 0 \text{ on } x_1 \leq x \leq t_q \text{ for} \\ &\quad \theta = 1, \dots, r_q; \quad q = 0, 1, \dots, g, \\ v_{ki}(x_1)v_{kj}(x_1) &= \delta_{ij} \quad (\delta_{ii} = 1, \delta_{ij} = 0 \text{ if } i \neq j) \quad (i, j = 1, \dots, n). \end{aligned}$$

Now define another set of n accessory extremals $u_{i \ n+j}(x), v_{i \ n+j}(x)$ by the initial conditions

$$(2.5) \quad u_{i \ n+j}(x_1) = v_{ij}(x_1), \quad v_{i \ n+j}(x_1) = 0 \quad (i, j = 1, \dots, n).$$

Finally, define $u_{ij}(x | q), v_{ij}(x | q)$ ($q = 0, 1, \dots, g$) as follows:

$$(2.6) \quad \begin{aligned} u_{ij}(x | q) &= u_{i \ n+j}(x), \quad v_{ij}(x | q) = v_{i \ n+j}(x) \text{ for } j = 1, \dots, r_q, \\ u_{ij}(x | q) &= u_{ij}(x), \quad v_{ij}(x | q) = v_{ij}(x) \text{ for } j = r_q + 1, \dots, n. \end{aligned}$$

The following property follows readily from the definition of a conjugate point:

PROPERTY 5°. A value x_3 on the sub-interval $t_{q+1} < x \leq t_q$ ($q = 0, 1, \dots, g$) defines a point on E_{12} conjugate to the point 1 if and only if one of the following conditions is satisfied:

(α) the matrix $\begin{vmatrix} u_{is}(x_1) \\ u_{is}(x_3) \end{vmatrix} \quad (s = 1, \dots, 2n)$ has rank less than $2n - r_q$.

(β) the matrix $\|u_{im}(x_3)\| \equiv \|u_{im}(x_3 | q)\|$ ($m = r_q + 1, \dots, n$) has rank less than $n - r_q$.

(γ) the determinant $|u_{ij}(x_3 | q)|$ is equal to zero.

It is verified readily that the accessory extremals of each of the n parameter families defined by (2.4), (2.5) or (2.6) are mutually conjugate [see I, p. 738] in pairs.

3. Proof of Theorem A. For clarity, the essential steps in the proof of this theorem will be stated in the form of lemmas.

LEMMA 3.1. In terms of the accessory extremals defined by (2.4) and (2.5), define n mutually conjugate accessory extremals $U_{ij}(x; \rho)$ $V_{ij}(x; \rho)$ as follows:

$$(3.1) \quad \begin{aligned} U_{ij}(x; \rho) &= u_{i, n+j}(x) + \rho u_{ij}(x); \quad V_{ij}(x; \rho) = v_{i, n+j}(x) + \rho v_{ij}(x), \\ &\quad \text{for } i = 1, \dots, n; \quad j = 1, \dots, r_g; \\ U_{ij}(x; \rho) &= u_{ij}(x), \quad V_{ij}(x; \rho) = v_{ij}(x) \quad \text{for } j = r_g + 1, \dots, n. \end{aligned}$$

Then if E_{12} is an extremal satisfying the conditions of Theorem A, the determinant $|U_{ij}(x; \rho)|$ is different from zero on $x_1 < x \leq x_2$ for ρ positive and sufficiently large in value.

This lemma will be proved by induction. In view of condition (γ) of Property 5° of § 2, it is seen that $|u_{ij}(x | q)| \neq 0$ on $t_{q+1} < x \leq t_q$ ($q = 0, 1, \dots, g$). Now for $x_1 < x \leq t_g$ we have $U_{ij}(x; \rho) \equiv u_{ij}(x | g)$, and hence for arbitrary values ρ we have $|U_{ij}(x; \rho)| \neq 0$ on $x_1 < x \leq t_g$.

It will now be proved that if for a value $q = \sigma$ there exists a positive value $\rho = \rho_1$ such that $|U_{ij}(x; \rho_1)| \neq 0$ on $x_1 < x \leq t_\sigma$, then there exists a value $\rho_2 > \rho_1$ such that $|U_{ij}(x; \rho_2)| \neq 0$ on $x_1 < x \leq t_{\sigma-1}$. By hypothesis, $|U_{ij}(x; \rho_1)| \neq 0$ on $x_1 < x \leq t_\sigma$. Hence there exists an ϵ such that $t_\sigma < t_\sigma + \epsilon < t_{\sigma-1}$ and $|U_{ij}(x; \rho_1)| \neq 0$ on $x_1 < x \leq t_\sigma + \epsilon$. It will first be shown that if $\rho > \rho_1$, then $|U_{ij}(x; \rho)| \neq 0$ also on $x_1 < x \leq t_\sigma + \epsilon$. For if there were a value x_3 on this interval such that $|U_{ij}(x_3; \rho)| = 0$, there would exist constants c_j not all zero and such that $U_{ij}(x_3; \rho)c_j = 0$. For these constants, let

$$(3.2) \quad \eta_i(x) = U_{ij}(x; \rho) c_j, \quad \xi_i(x) = V_{ij}(x; \rho) c_j \quad (x_1 \leq x \leq x_3).$$

Since on $x_1 t_g$ we have $U_{ij}(x; \rho) \equiv U_{ij}(x; \rho_1)$, and $|U_{ij}(x; \rho_1)| \neq 0$ on $x_1 < x \leq t_\sigma + \epsilon$, it would follow that the functions $\eta_i(x)$ defined by (3.2) are of the form $\eta_i(x) = U_{ij}(x; \rho_1) a_j(x)$ on $x_1 \leq x \leq x_3$, and the functions $a_j(x)$ are of class C' on this interval. Moreover, on $x_1 t_g$ we have $a_j(x) = c_j$ ($j = 1, \dots, n$). In view of condition III' and the Clebsch transformation of the second variation [I, p. 739], it would follow that

$$(3.3) \quad \int_{x_1}^{x_3} 2\omega[x, \eta, \eta'] dx + \eta_i(x_1) V_{ij}(x_1; \rho_1) c_j \\ = \int_{x_1}^{x_3} F_{y' i y' j} \{U_{ik}(x; \rho_1) a'_k(x)\} \{U_{jl}(x; \rho_1) a'_l(x)\} dx \geq 0.$$

On the other hand, by direct integration we obtain

$$\int_{x_1}^{x_3} 2\omega[x, \eta, \eta'] dx = -\eta_i(x_1) \xi_i(x_1),$$

and as a consequence of the initial values of $U_{ij}(x; \rho)$ and $V_{ij}(x; \rho)$, it would follow that

$$(3.4) \quad \int_{x_1}^{x_3} 2\omega[x, \eta, \eta'] dx + \eta_i(x_1) V_{ij}(x_1; \rho_1) c_j = -(\rho - \rho_1) \sum_{j=1}^{r_g} c_j^2.$$

Relation (3.4) is seen to be a contradiction to (3.3) unless $c_\tau = 0$ ($\tau = 1, \dots, r_g$). In this latter case, since $|U_{ij}(x_3; \rho_1)| \neq 0$ it would follow that $c_j = 0$ ($j = 1, \dots, n$), which is a contradiction. We have proved, therefore, that if $|U_{ij}(x; \rho_1)| \neq 0$ on $x_1 < x \leq t_\sigma + \epsilon$, then for $\rho > \rho_1$, we have $|U_{ij}(x; \rho)| \neq 0$ on this interval.

Finally, it is to be noted that on $t_\sigma + \epsilon \leq x \leq t_{\sigma-1}$ the determinant $|U_{ij}(x; \rho)|$ is a polynomial in ρ of degree $r_g - r_{\sigma-1}$ whose leading coefficient is $|u_{ij}(x; \sigma - 1)|$, and therefore different from zero. Hence for ρ sufficiently large in absolute value we have $|U_{ij}(x; \rho)| \neq 0$ on $t_\sigma + \epsilon \leq x \leq t_{\sigma-1}$. Combining these results, we have that there exists a positive value ρ_2 such that $\rho_2 > \rho_1$ and $|U_{ij}(x; \rho_2)| \neq 0$ on $x_1 < x \leq t_{\sigma-1}$. We have established, therefore, an induction proof of Lemma 3.1.

In order to complete the proof of Theorem A, let $x_0 = x_1 - (t_g - x_1)$, and define the coefficients of ω and Φ_a on $x_0 \leq x < x_1$ by the following identities in (x, η, η') :

$$(3.5) \quad \omega[x, \eta, \eta'] \equiv \omega[2x_1 - x, -\eta, \eta'], \quad \Phi_a[x, \eta, \eta'] \equiv \Phi_a[2x_1 - x, -\eta, \eta'].$$

In the canonical system (2.3') we then have for $x_0 \leq x < x_1$,

$$(3.6) \quad \begin{aligned} A_{ij}(x) &= -A_{ij}(2x_1 - x), & B_{ij}(x) &= B_{ij}(2x_1 - x), \\ C_{ij}(x) &= C_{ij}(2x_1 - x). \end{aligned}$$

If $\eta_i(x) \equiv 0$, $\xi_i(x)$ is a solution of (2.3') on x_0x_1 , then $\eta_i(x) \equiv 0$, $\xi_i(2x_1 - x)$ is a solution of this system on x_1t_g , and conversely. Hence on every sub-interval x_3x_1 of x_0x_1 the order of anormality is r_g . Moreover, on x_0x_1 , the condition III' holds.

It is to be emphasized that after the coefficients of the system (2.3), or (2.3'), have been defined on x_0x_1 by relations (3.5), these coefficients will in general be discontinuous at $x = x_1$. Hence, by a solution of this system on x_0x_2 we shall understand a set of functions $\eta_i(x)$, $\xi_i(x)$ which are continuous on this interval, and which, except possibly at x_1 , have derivatives satisfying equations (2.3'). We shall still denote by $u_{ij}(x)$, $v_{ij}(x)$, $x_0 \leq x \leq x_2$, the solutions of this modified system on x_0x_2 which satisfy the initial conditions (2.4). A set of r_g linearly independent solutions with $\eta_i(x) \equiv 0$ on x_0x_1 is given by the functions $u_{i\tau}(x)$, $v_{i\tau}(x)$ ($\tau = 1, \dots, r_g$).

Now let $v^*_{ij}(x)$ ($j = 1, \dots, n$) be continuous functions such that $v^*_{ij}(x_1) = v_{ij}(x_1)$, and

$$(3.7) \quad v_{i\tau}(x)v^*_{ij}(x) \equiv 0 \text{ on } x_0x_1 \text{ if } \tau \neq j \text{ } (\tau = 1, \dots, r_g; j = 1, \dots, n).$$

Moreover, let ρ be such that $|U_{ij}(x; \rho)| \neq 0$ on $x_1 < x \leq x_2$.

Corresponding to a point t on x_0x_1 there exist mutually conjugate solutions $\eta_{ij}(x; t)$, $\xi_{ij}(x; t)$ of (2.3') on x_0x_2 such that

$$(3.8) \quad \begin{aligned} \eta_{ij}(t; t) &= v_{ij}(t), & \xi_{ij}(t; t) &= \rho v^*_{ij}(t) \text{ for } (j = 1, \dots, r_g), \\ \eta_{ij}(t; t) &= 0, & \xi_{ij}(t; t) &= v^*_{ij}(t) \text{ for } (j = r_g + 1, \dots, n). \end{aligned}$$

For $t = x_1$ the initial conditions (3.8) reduce to the initial values of $U_{ij}(x; \rho)$, $V_{ij}(x; \rho)$. Hence, $\eta_{ij}(x; x_1) \equiv U_{ij}(x; \rho)$, $\xi_{ij}(x; x_1) \equiv V_{ij}(x; \rho)$ on x_0x_2 .

It will now be shown that for $t < x_1$ and sufficiently near to x_1 the determinant $|\eta_{ij}(x; t)| \neq 0$ on x_1x_2 . It will first be noted that on a sub-interval $x'x''$, where $x_0 \leq x' < x_1$ and $x_1 \leq x'' \leq t_g$, the order of anormality is r_g . The following lemma may then be proved by the method used by Bliss [I, p. 739] to prove a corresponding result for an arc which satisfies stronger normality conditions.

LEMMA 3.2. *There exists an interval $x_1 - d \leq x \leq x_1 + d$ ($0 < d < t_g - x_1$) such that if $\eta_i(x)$, $\xi_i(x)$ is a solution of (2.3'), and the functions $\eta_i(x)$ all vanish at a point x' of $x_1 - d \leq x < x_1$ and at a*

point x'' of $x_1 \leq x \leq x_1 + d$, then on $x'x''$ we have $\eta_i(x) \equiv 0$, $\xi_i(x) = v_{i\tau}(x)c_\tau$, where c_τ ($\tau = 1, \dots, r_g$) are constants.

From the initial values of the functions $v^*_{ij}(x)$ it is seen that there exists a δ such that $0 < \delta < d$, where d is determined as in Lemma 3.2, and such that if $x_1 - \delta \leq x \leq x_1$, then the determinant $|v_{i\tau}(x)v^*_{i\sigma}(x)|$ ($\tau = 1, \dots, r_g$; $\sigma = r_g + 1, \dots, n$) is different from zero. It then follows, from the result of Lemma 3.2, that if t is an arbitrary point on $x_1 - \delta \leq t < x_1$ and $\eta_{ij}(x; t)$, $\xi_{ij}(x; t)$ is the corresponding family determined by the initial conditions (3.8), then $|\eta_{ij}(x; t)| \neq 0$ on $x_1 \leq x \leq x_1 + d$.

Finally, since $\eta_{ij}(x; x_1) \equiv U_{ij}(x; \rho)$, it follows from the continuity of the solutions $\eta_{ij}(x; t)$, $\xi_{ij}(x; t)$ of (2.3') when considered as functions of t , that for t sufficiently near to x_1 the determinant $|\eta_{ij}(x; t)|$ is also different from zero on $x_1 + d \leq x \leq x_2$. Therefore, for $t < x_1$ and sufficiently near to x_1 the family of mutually conjugate solutions of (2.3') determined by the conditions (3.8) is such that $|\eta_{ij}(x; t)| \neq 0$ on $x_1 \leq x \leq x_2$. This completes the proof of Theorem A.

The following corollary is an immediate consequence of well-known results for the problem of Lagrange:

COROLLARY. Suppose E_{12} is an extremal arc which satisfies the conditions of Theorem A. If $u_i(x)$, $v_i(x)$ is an accessory extremal, and $\eta_i(x)$ is an arbitrary admissible variation such that $\eta_i(x_1) = u_i(x_1)$, $\eta_i(x_2) = u_i(x_2)$, then

$$(3.9) \quad \int_{x_1}^{x_2} 2\omega[x, \eta, \eta'] dx \geq \int_{x_1}^{x_2} 2\omega[x, u, u'] dx,$$

and the equality sign holds if and only if $\eta_i(x) \equiv u_i(x)$ on x_1x_2 .

4. Further discussion of Theorem A. It has been established in Lemma 3.1 that along an extremal arc which satisfies the conditions of Theorem A the accessory extremals $U_{ij}(x; \rho)$, $V_{ij}(x; \rho)$ defined by (3.1) are such that $|U_{ij}(x; \rho)| \neq 0$ on $x_1 < x \leq x_2$ for $\rho > 0$ and sufficiently large in value. If $r_g = r_0$, that is, if the order of anormality of E_{12} is the same on every sub-interval x_1x of x_1x_2 , it is seen from the first paragraph of the proof of Lemma 3.1 that this condition is true for arbitrary values of ρ . For further discussion we shall assume $r_g > r_0$, and seek to determine a lower bound for the values of ρ which are such that the conclusion of Lemma 3.1 is satisfied. This lower bound is not determined independent of the results of § 3, however, since use is made of the Corollary of that section.

In view of condition IV' and condition (α) of property 5° in § 2, it is

seen that the $2n - r_0$ accessory extremals $u_{i\kappa}(x)$, $v_{i\kappa}(x)$ ($\kappa = r_0 + 1, \dots, 2n$) defined by (2.4) and (2.5) are such that the matrix

$$\begin{vmatrix} u_{i\kappa}(x_1) \\ u_{i\kappa}(x_2) \end{vmatrix}$$

is of rank $2n - r_0$. As a consequence of property 2° of § 2, or by a direct consideration of the initial values $u_{is}(x_1)$ ($s = 1, \dots, 2n$), it is seen that the matrix of $2n - r_g$ rows and $2n - r_0$ columns

$$(4.1) \quad \begin{vmatrix} v_{i\sigma}(x_1)u_{i\kappa}(x_1) \\ u_{i\kappa}(x_2) \end{vmatrix} \quad \begin{matrix} (\sigma = r_g + 1, \dots, n; i = 1, \dots, n; \\ \kappa = r_0 + 1, \dots, 2n) \end{matrix}$$

is of rank $2n - r_g$.

For simplicity of notation, let $2A[z] \equiv a_{\kappa\lambda} z_\kappa z_\lambda$ ($\kappa, \lambda = r_0 + 1, \dots, 2n$) denote the quadratic form

$$(4.2) \quad \int_{x_1}^{x_2} 2\omega[x, u_{i\kappa} z_\kappa, u'_{i\kappa} z_\kappa] dx = z_\kappa u_{i\kappa}(x) v_{i\lambda}(x) z_\lambda \Big|_{x=x_1}^{x=x_2}.$$

We shall also denote by $B_\nu[z] \equiv b_{\nu\kappa} z_\kappa$ ($\nu = 1, \dots, 2n - r_g$) the first members of the linear equations

$$(4.3) \quad \begin{matrix} v_{i\sigma}(x_1)u_{i\kappa}(x_1)z_\kappa = 0, & (\sigma = r_g + 1, \dots, n; i = 1, \dots, n; \\ u_{i\kappa}(x_2)z_\kappa = 0, & \kappa = r_0 + 1, \dots, 2n) \end{matrix}$$

Finally, denote by $2K[z] \equiv k_{\kappa\lambda} z_\kappa z_\lambda$ the quadratic form

$$(4.4) \quad [u_{i\kappa}(x_1)u_{i\lambda}(x_1) + u_{i\kappa}(x_2)u_{i\lambda}(x_2)]z_\kappa z_\lambda,$$

which, in view of IV' and elementary properties of matrices, is positive definite. It then follows that the class of values (z_κ) which satisfy the conditions

$$(4.5) \quad B_\nu[z] = 0, \quad 2K[z] = 1$$

is not vacuous. Moreover, the minimum value of $2A[z]$ in this class of values is the smallest zero l_1 of the determinant⁹

⁹ See, for example, Hancock, *Theory of Maxima and Minima*, Ginn and Co., Boston (1917), pp. 103-114. Bliss has phrased his analogue of the Jacobi condition for the problem of Bolza in terms of the roots of a determinant of the form (4.6); see [IV], p. 273. It may be shown that corresponding to each of the zeros $l = l_h$ ($h = 1, \dots, r_g - r_0$) of $D(l)$ there exists an accessory extremal $\eta_{ih}(x)$, $\xi_{ih}(x)$ satisfying with constants $d_{\sigma h}$ the end conditions:

$$v_{i\sigma}(x_1)\eta_{ih}(x_1) = 0, \quad \eta_{ih}(x_2) = 0; \quad d_{\sigma h}v_{i\sigma}(x_1) - l_h\eta_{ih}(x_1) - \xi_{ih}(x_1) = 0.$$

See [IX], § 5, and also [V].

$$(4.6) \quad D(l) = \begin{vmatrix} a_{\kappa\lambda} - lk_{\kappa\lambda} & b_{v\kappa} \\ b_{v\lambda} & 0 \end{vmatrix}.$$

Finally, in view of IV'_0 we have that if $\eta_i(x)$ is an arbitrary admissible variation, then there exists a unique accessory extremal $u_i = u_{i\kappa}(x)z_\kappa$, $v_i = v_{i\kappa}(x)z_\kappa$, such that $u_i(x_1) = \eta_i(x_1)$, $u_i(x_2) = \eta_i(x_2)$ [X, p. 809]. As a consequence of the minimizing property of l_1 and the corollary of § 3, we have that if $\eta_i(x)$ is an admissible variation such that

$$(4.7) \quad v_{i\sigma}(x_1)\eta_i(x_1) = 0, \quad \eta_i(x_2) = 0, \quad \eta_i(x_1)\eta_i(x_1) \neq 0,$$

then
$$\int_{x_1}^{x_2} 2\omega[x, \eta, \eta'] dx \geq l_1[\eta_i(x_1)\eta_i(x_1)].$$

Now suppose that for a given value of ρ there exists a point x_3 such that $x_1 < x_3 \leq x_2$ and $|U_{ij}(x_3; \rho)| = 0$. If c_j ($j = 1, \dots, n$) are constants not all zero and such that $U_{ij}(x_3; \rho)c_j = 0$, it follows from IV'_0 that $c_\tau c_\tau \neq 0$, ($\tau = 1, \dots, r_g$). Moreover, the admissible arc defined by

$$\eta_i(x) = U_{ij}(x; \rho)c_j \text{ on } x_1x_3; \quad \eta_i(x) \equiv 0 \text{ on } x_3x_2,$$

satisfies equations (4.7). On direct integration, we obtain

$$\int_{x_1}^{x_2} 2\omega[x, \eta, \eta'] dx = -\rho c_\tau c_\tau = -\rho[\eta_i(x_1)\eta_i(x_1)],$$

and we must have, therefore, $\rho \leq -l_1$. We have established, therefore, the following theorem:

THEOREM 4.1. *Suppose that E_{12} is an extremal arc satisfying the conditions of Theorem A, and $U_{ij}(x; \rho)$, $V_{ij}(x; \rho)$ is the system of accessory extremals defined by (3.1). Then for $\rho > -l_1$, where l_1 is the smallest zero of the determinant $D(l)$, we have $|U_{ij}(x; \rho)| \neq 0$ on $x_1 < x \leq x_2$.*

Finally, there will be given a method for determining a system of mutually conjugate accessory extremals $\eta_{ij}(x)$, $\xi_{ij}(x)$ with $|\eta_{ij}(x)| \neq 0$ on x_1x_2 which does not use directly the system defined by (3.1).

Consider the problem of minimizing $2A[z]$ in the class of values (z_κ) ($\kappa = r_0 + 1, \dots, 2n$) which satisfy the conditions

$$(4.8) \quad u_{i\kappa}(x_2)z_\kappa = 0, \quad 2K[z] = 1,$$

where $A[z]$ and $K[z]$ are defined by (4.2) and (4.4). The minimum value of $2A[z]$ in this class of values is the smallest zero m_1 of the determinant

$$(4.9) \quad \mathfrak{D}(m) = \begin{vmatrix} a_{\kappa\lambda} - m k_{\kappa\lambda} & u_{j\kappa}(x_2) \\ u_{i\lambda}(x_2) & 0 \end{vmatrix}.$$

By the use of the Corollary of § 3, and by an argument like that used above, it is seen that if $\eta_i(x)$ is an admissible variation such that $\eta_i(x_2) = 0$, $\eta_i(x_1)\eta_i(x_1) \neq 0$, then

$$\int_{x_1}^{x_2} 2\omega[x, \eta, \eta'] dx \geq m_1[\eta_i(x_1)\eta_i(x_1)].$$

In terms of the accessory extremals (2.4) and (2.5), now define a system of mutually conjugate accessory extremals $Y_{ij}(x; \rho)$, $Z_{ij}(x; \rho)$ as follows

$$(4.10) \quad Y_{ij}(x; \rho) = u_{i, n+j}(x) + \rho u_{ij}(x), \quad Z_{ij}(x; \rho) = v_{i, n+j}(x) + \rho v_{ij}(x).$$

Suppose that for a given value of ρ there exists a point x_3 such that $x_1 < x_3 \leq x_2$ and $|Y_{ij}(x_3; \rho)| = 0$. Let d_j ($j = 1, \dots, n$) be constants not all zero such that $Y_{ij}(x_3; \rho)d_j = 0$. If we define

$$\eta_i(x) = Y_{ij}(x; \rho)d_j \text{ on } x_1x_3, \quad \eta_i(x) \equiv 0 \text{ on } x_3x_2,$$

then $\eta_i(x_1)\eta_i(x_1) = d_id_i \neq 0$. Moreover, by direct integration we obtain

$$\int_{x_1}^{x_2} 2\omega[x, \eta, \eta'] dx = -\rho d_id_i = -\rho[\eta_i(x_1)\eta_i(x_1)].$$

Hence, if ρ is a value such that $|Y_{ij}(x; \rho)| = 0$ at a point x_3 on $x_1 < x \leq x_2$, it follows that $\rho \leq -m_1$. Finally, it is to be noted that

$$|Y_{ij}(x_1; \rho)| = |u_{i, n+j}(x_1)| \neq 0.$$

We have, therefore, the following theorem.

THEOREM 4.2. *Suppose that E_{12} is an extremal arc satisfying the conditions of Theorem A, and that $Y_{ij}(x; \rho)$, $Z_{ij}(x; \rho)$ is the family of mutually conjugate accessory extremals defined by (4.10). Then for $\rho > -m_1$, where m_1 is the smallest zero of the determinant $\mathfrak{D}(m)$, we have $|Y_{ij}(x; \rho)| \neq 0$ on $x_1 \leq x \leq x_2$.*

In Theorem 4.2 we have assumed that E_{12} is an extremal satisfying the conditions of Theorem A, that is, it satisfies the strengthened Clebsch condition and has on it no point conjugate to the point 1. Consequently, in the proof of Theorem 4.2 use is made of the Corollary of § 3. It is significant to note, however, that quite independent of the results of § 3 the above proof of Theorem 4.2 leads to a result which is of itself important.

In the proof of the above theorem explicit use is made of only the following hypotheses: (1) E_{12} is an extremal arc satisfying the strengthened Clebsch condition, (2) the point 2 is not conjugate to the point 1, and (3) if $u_i(x)$, $v_i(x)$ is an accessory extremal and $\eta_i(x)$ is an arbitrary admissible variation such that $\eta_i(x_1) = u_i(x_1)$, $\eta_i(x_2) = u_i(x_2)$, then the inequality (3.9) holds. Let μ_a denote the multipliers corresponding to the accessory extremal u_i, v_i . Then by an expansion of the integrand function $2\omega[x, \eta, \eta'] = 2\Omega[x, \eta, \eta', \mu]$ by Taylor's formula, and an integration by parts, one obtains that assumption (3) is equivalent to the condition that if $\eta_i(x)$ is an arbitrary admissible variation such that $\eta_i(x_1) = 0 = \eta_i(x_2)$, then

$$(4.10) \quad \int_{x_1}^{x_2} 2\omega[x, \eta, \eta'] dx \geq 0.$$

In view of this remark we have, therefore, that quite independent of § 3 the above proof of Theorem 4.2 leads to the following result:

THEOREM 4.3. *Suppose E_{12} is an extremal arc which satisfies the strengthened Clebsch condition, on which the point 2 is not conjugate to the point 1, and such that if $\eta_i(x)$ is an arbitrary admissible variation along E_{12} having $\eta_i(x_1) = 0 = \eta_i(x_2)$, then inequality (4.10) is satisfied. Then for $\rho > -m_1$, where m_1 is the smallest zero of the determinant $\mathfrak{D}(m)$, we have $|Y_{ij}(x; \rho)| \neq 0$ on $x_1 \leq x \leq x_2$.*

It is to be remarked that one may show the equivalence of the hypotheses of Theorems 4.2 and 4.3 by the use of the results of § 3 and the first necessary condition for the problem of Lagrange. However, the significance of Theorem 4.3 lies in its simplicity of proof and its usefulness. For example, with the aid of this theorem one may establish in a simple manner the existence of infinitely many characteristic numbers for the boundary value problem which the author has previously treated by the use of other methods [see V and IX].

5. Relation of Theorem A to sufficiency theorems for the problem of Bolza. By the use of Theorem A the proof of a sufficiency theorem for the problem of Bolza as given by Bliss [IV] is still valid when the normality conditions there used are replaced by the weaker assumptions of normality with respect to the end conditions and normality on x_1x_2 . This latter assumption of normality is used by Bliss not only in the proof of the imbedding theorem, but also in the proof of another result [IV, Lemma 2, p. 267]. Hence, Theorem A does not eliminate entirely the assumption of normality

on x_1x_2 in the proof of sufficient conditions as given by Bliss. The result of Theorem A also enables one to establish sufficiency theorems for the problem of Mayer by the methods used by Bliss and Hestenes [VI] and Hestenes [VII] under correspondingly weakened hypotheses.

In a recent paper [IX] the author has discussed for the problem of Mayer the relations between the boundary value problem formulation of the analogue of Jacobi's condition and the analogue of this condition introduced by Bliss {see [IV] and [VII]}. As a consequence of Theorem A the results of Theorems 4.1, 4.2 and 4.4 of that paper are valid when the assumptions concerning normality on sub-intervals is omitted. Consequently, these results apply directly to the problem of Bolza as well as to the problem of Mayer. In particular, using the terminology of that paper, we have:

If E_{12} is an extremal of the form $y_i = y_i(x)$, $\lambda_0 = 1$, $\lambda_a(x)$ which satisfies conditions I and III', then condition IV'_C is satisfied if and only if conditions IV'_0 and IV'_B are satisfied.

For the problem of Bolza as treated by Hestenes [X], Theorem A leads to the following results, when expressed in Hestenes' notation [X, pp. 810-811].

If g is an extremal satisfying conditions I and III', then condition V' is satisfied if and only if condition VI' is satisfied along g . Again, using the notation of Hestenes, we have the following relation between the conditions IV'_C , IV'_0 and IV'_B of [IX], and conditions V' and VI' of Hestenes {see [X], pp. 810-816}.

Suppose that the end conditions are regular, and that the non-tangency condition holds on an admissible arc g having no corners. If g satisfies conditions I, III' with a set of multipliers $\lambda_0 = 1$, $\lambda_\beta(x)$, then each of the four following conditions implies the others: IV'_C , IV'_0 with IV'_B , V' , VI'. In case the end conditions are separated, then each of the above conditions is equivalent to the condition IV' of Hestenes [X, p. 806].

The above results lead to obvious simplification of the sufficiency theorems as stated by Hestenes.

BIBLIOGRAPHY.

I. Bliss, "The problem of Lagrange in the calculus of variations," *American Journal of Mathematics*, vol. 52 (1930), pp. 673-747.

II. Morse, "Sufficient conditions in the problem of Lagrange with fixed end points," *Annals of Mathematics*, ser. 2, vol. 32 (1931), pp. 567-577.

III. Morse, "Sufficient conditions in the problem of Lagrange with variable end conditions," *American Journal of Mathematics*, vol. 53 (1931), pp. 517-546.

IV. Bliss, "The problem of Bolza in the calculus of variations," *Annals of Mathematics*, ser. 2, vol. 33 (1932), pp. 261-274.

V. Reid, "A boundary value problem associated with the calculus of variations," *American Journal of Mathematics*, vol. 54 (1932), pp. 769-790.

VI. Bliss and Hestenes, "Sufficient conditions for a problem of Mayer in the calculus of variations," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 305-326; *Contributions to the Calculus of Variations 1931-1932*, The University of Chicago Press, pp. 295-337.

VII. Hestenes, "Sufficient conditions for the general problem of Mayer with variable end points," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 479-490; *Contributions to the Calculus of Variations 1931-1932*, The University of Chicago Press, pp. 339-360.

VIII. Hu, "The problem of Bolza and its accessory boundary value problem," *Contributions to the Calculus of Variations 1931-1932*, The University of Chicago Press, pp. 361-443.

IX. Reid, "Analogues of the Jacobi condition for the problem of Mayer in the calculus of variations," *Annals of Mathematics*, ser. 2, vol. 35 (1934), pp. 836-848.

X. Hestenes, "Sufficient conditions for the problem of Bolza in the calculus of variations," *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 793-818.

THE UNIVERSITY OF CHICAGO,
CHICAGO, ILLINOIS.

CONCERNING SOME METHODS OF BEST APPROXIMATION, AND A THEOREM OF BIRKHOFF.¹

By I. M. SHEFFER.

Introduction. The series of Taylor has been generalized in many directions. It is our purpose here to consider a natural extension of certain series in addition to that of Taylor, by methods analogous to those used by Birkhoff² and by Widder³ for Taylor series.

The Widder theory is based on a best approximation definition. Let $\{\phi_n(x)\}$, $(n = 0, 1, \dots)$ be an infinite sequence of functions, analytic about $x = 0$. An expression of the form $s_n(x) = \sum_0^n c_i \phi_i(x)$ is a "polynomial" of order (not exceeding) n . According to Widder, the "polynomial" $s_n(x)$ is the best approximation of order n to the function $f(x)$, analytic about $x = 0$, if $f(x) - s_n(x)$ vanishes, together with its first n derivatives, at $x = 0$. If certain determinants do not vanish, then s_n exists and is unique. The question of convergence of $s_n(x)$ to $f(x)$ is equivalent to that of $s_0(x) + \sum_1^\infty [s_n(x) - s_{n-1}(x)]$ to $f(x)$. Now a striking situation comes to light: $s_n - s_{n-1}$ can be written in the form

$$s_n(x) - s_{n-1}(x) = f_n \Omega_n(x)$$

where $\Omega_n(x)$ is a "polynomial" of order n , independent of $f(x)$, the function $f(x)$ making its presence felt only in the constants $\{f_n\}$. That is, $s_0(x) + \sum_1^\infty [s_n(x) - s_{n-1}(x)] \sim \sum_0^\infty f_n \Omega_n(x)$; so that in going from the n -th approximation to the $(n+1)$ -st, we have only to add in the term $f_{n+1} \Omega_{n+1}(x)$, the terms already present remaining. Such behavior (which is also true of the extension we have in view) we shall refer to as the *property of permanence*.

In the Widder case,

$$\Omega_n(x) = (x^n/n!) [1 + h_n(x)],$$

¹ Presented to the American Mathematical Society, December, 1933.

² G. D. Birkhoff, "Sur une généralisation de la série de Taylor," *Comptes Rendus*, vol. 164 (1917), pp. 942-945.

³ D. V. Widder, "On the expansion of analytic functions of the complex variable in generalized Taylor's series," *Transactions of the American Mathematical Society*, vol. 31 (1929), pp. 43-52. (This paper contains a reference to a preceding work by Widder in which functions of a *real variable* are considered; but that phase of the problem does not concern us here.)

where $h_n(x)$ is analytic at $x=0$ and $h_n(0)=0$; and his convergence theory is based on the following additional hypotheses:

- (i) $h_n(x)$ is analytic, $|x| \leq R$;
- (ii) M exists, independent of n and x , such that $|h_n(x)| \leq M/(n+1)$, $|x| \leq R$.

In § 1 we give a rather general definition of best approximation, and establish the property of permanence. As particular cases are the Widder case, and the "least square" case. In § 2 we show that condition (ii) of Widder can be replaced by the less restrictive condition:

- (ii') $1 + h_n(x)$ converges uniformly, in $|x| \leq R$, to a function $M(x)$, with $M(x) \neq 0$ in $|x| \leq R$.

Essentially, we replace Widder's condition $|h_n(x)| = O(1/n)$ by the condition $|h_n(x)| = o(1)$. Finally, in § 3, to handle convergence in some other cases, we appeal to the method used by Birkhoff, which by means of an integral equation extends the convergence properties of Taylor series (i. e., series in $\{x^n\}$) to series in $\{v_n(x)\}$, where $\{v_n(x)\}$ is "sufficiently close" to $\{x^n\}$. Only, the rôle of $\{x^n\}$ is now played by functions $\{u_n(x)\}$, which we endow with properties analogous to those of the known functions $\{x^n\}$.

1. *Some methods of best approximation.* Let $\{\phi_n(x)\}$, ($n=0, 1, \dots$) be a sequence of functions. We wish to assign to each function $f(x)$ (of a certain class) a sequence of "polynomials" $\{s_n(x)\}$:

$$(1) \quad s_n(x) = c_{n0}\phi_0(x) + \dots + c_{nn}\phi_n(x),$$

which are the *best* approximating "polynomials" to $f(x)$, each of its order. This requires that we define the test for best approximation.

Let L_n , ($n=0, 1, \dots$) be a sequence of linear operators, each of which assigns to a function $u(x)$ a number: $L_n[u(x)] = u_n$.

DEFINITION. By the method \mathcal{M} of best approximation, relative to the set of operators L_n , is meant that determination of the set $\{s_n(x)\}$ according to the following test of best approximation:⁴

$$(2) \quad L_i[s_n(x)] = L_i[f(x)], \quad (i=0, 1, \dots, n).$$

⁴The "polynomials" $s_n(x)$ depend, of course, on the sequence of functions $\{\phi_n(x)\}$.

DEFINITION. A method \mathcal{M} is non-singular, relative to a sequence $\{\phi_n(x)\}$, if none of the following determinants vanishes:⁵

$$(3) \quad \Delta_n = \begin{vmatrix} L_0[\phi_0] & L_0[\phi_1] & \cdots & L_0[\phi_n] \\ L_1[\phi_0] & L_1[\phi_1] & \cdots & L_1[\phi_n] \\ \cdot & \cdot & \cdot & \cdot \\ L_n[\phi_0] & L_n[\phi_1] & \cdots & L_n[\phi_n] \end{vmatrix}, \quad (n=0, 1, \cdots).$$

We shall consider only non-singular methods.

THEOREM 1. For a given function⁶ $f(x)$, to each $(n=0, 1, \cdots)$ there is a unique best approximating "polynomial" $s_n(x)$ of order (not greater than) n .

This follows from equations (2), since $s_n(x)$ has the form (1), so that the determinant of system (2) is precisely $\Delta_n \neq 0$.

Let us form from $\{\phi_n(x)\}$ a new sequence $\{\Phi_n(x)\}$, linearly dependent on $\{\phi_n(x)\}$:

$$(4) \quad \Phi_n(x) = b_{n0}\phi_0(x) + \cdots + b_{nn}\phi_n(x), \quad b_{nn} \neq 0.$$

Clearly,⁷ the set of best approximating "polynomials" for $\{\Phi_n(x)\}$ coincides with the set $\{s_n(x)\}$ already found for $\{\phi_n(x)\}$. We may therefore choose one set out of the infinite number of possible ones, to be obtained from $\{\phi_n(x)\}$ as was $\{\Phi_n(x)\}$, to represent all such. There exists a "significant" set, which we shall term the basic set.

DEFINITION. The basic set for a given sequence $\{\phi_n(x)\}$, relative to a method \mathcal{M} , is the set $\{\Phi_n(x)\}$ defined by

$$(5) \quad L_i[\Phi_n(x)] = \begin{cases} 0, & (i=0, 1, \cdots, n-1); \\ 1, & (i=n), \end{cases}$$

where $\Phi_n(x)$ has the form (4).

In virtue of the condition $\Delta_n \neq 0$, we see that⁸ a basic set exists and is unique.

⁵ A method \mathcal{M} may be essentially singular; i.e., is singular for all sequences $\{\phi_n(x)\}$, as for example if n exists such that L_0, L_1, \cdots, L_n are linearly dependent. On the other hand, \mathcal{M} may be in general non-singular, but a peculiar choice of $\{\phi_n(x)\}$ may give singularity, as for example if n exists such that ϕ_0, \cdots, ϕ_n are linearly dependent.

⁶ It is understood that $f(x)$ and the functions $\phi_n(x)$ are within the class of functions on which the L_n 's can operate.

⁷ The determinants Δ_n for $\{\Phi_n(x)\}$ are non-vanishing, as is easily seen from (3).

⁸ The only point not obvious is that in $\Phi_n(x)$, $b_{nn} \neq 0$. But if $b_{nn} = 0$, then the

LEMMA 1. *There is a sequence of constants $\{f_n\}$ such that*

$$(6) \quad s_0(x) = f_0 \Phi_0(x); \quad s_n(x) - s_{n-1}(x) = f_n \Phi_n(x), \quad (n = 1, 2, \dots).$$

Since s_0 and Φ_0 are each multiples of ϕ_0 , and $b_{00} \neq 0$, therefore f_0 can be found. Now consider $T_n = s_n - s_{n-1}$. We have from (2):

$$L_i[T_n] = 0, \quad (i = 0, 1, \dots, n-1); \quad L_n[T_n] = f_n,$$

where we set

$$(7) \quad f_n = L_n[s_n(x) - s_{n-1}(x)].$$

If $f_n = 0$, then since $\Delta_n \neq 0$ we have $T_n(x) \equiv 0$, thus satisfying (6). If $f_n \neq 0$, then T_n/f_n satisfies equations (5), whence by uniqueness, $T_n/f_n \equiv \Phi_n$, and again (6) holds.

COROLLARY 1. *Method \mathcal{M} has the permanence property:⁹*

$$(8) \quad s_n(x) = s_0(x) + \sum_{i=1}^n [s_i(x) - s_{i-1}(x)] = \sum_{i=0}^n f_i \Phi_i(x),$$

where only the constants f_i (which are independent of n) depend on the function $f(x)$.

COROLLARY 2. *The constants $\{f_n\}$ are given by*

$$(9) \quad f_n = \frac{1}{\Delta_{n-1}} \begin{vmatrix} L_0[\phi_0] & L_0[\phi_1] & \cdots & L_0[\phi_{n-1}] & L_0[f] \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ L_n[\phi_0] & L_n[\phi_1] & \cdots & L_n[\phi_{n-1}] & L_n[f] \end{vmatrix},$$

$$(n = 1, 2, \dots); \quad f_0 = L_0[f].$$

For, let g_n denote the determinant in the numerator of the right-hand side of (9). In (6), $s_n - s_{n-1}$ and $f_n \Phi_n$ are linear combinations of the functions ϕ_0, \dots, ϕ_n which are (as we have observed in a footnote) linearly independent. Hence coefficients of ϕ_n on both sides of (6) must be equal. From (2) this coefficient in $s_n - s_{n-1}$ is g_n/Δ_n , and from (5) this coefficient in $f_n \Phi_n$ is $f_n \Delta_{n-1}/\Delta_n$. Hence $f_n = g_n/\Delta_{n-1}$, which is (9).

COROLLARY 3. *The functions $\{\Phi_n(x)\}$ are given by*

first n equations of (5) tell us, since $\Delta_{n-1} \neq 0$, that $b_{n0} = b_{n1} = \dots = b_{n,n-1} = 0$, so that the $(n+1)$ -st equation of (5) is not satisfied; a contradiction.

⁹ From this follows the curious fact that if we choose $f(x) = \Phi_n(x)$, then zero is the best approximating "polynomial" of all orders less than n .

$$(10) \quad \Phi_n(x) = \frac{1}{\Delta_n} \begin{vmatrix} L_0[\phi_0] & \cdots & L_0[\phi_n] \\ \vdots & \ddots & \vdots \\ L_{n-1}[\phi_0] & \cdots & L_{n-1}[\phi_n] \\ \phi_0(x) & \cdots & \phi_n(x) \end{vmatrix},$$

$$(n = 1, 2, \cdots); \quad \Phi_0(x) = \phi_0(x)/\Delta_0.$$

(10) follows at once from (5) and the uniqueness of a basic set.

COROLLARY 4. The "polynomials" $s_n(x)$ are given by

$$(11) \quad s_n(x) = \frac{-1}{\Delta_n} \begin{vmatrix} L_0[\phi_0] & \cdots & L_0[\phi_n] & L_0[f] \\ \vdots & \ddots & \vdots & \vdots \\ L_n[\phi_0] & \cdots & L_n[\phi_n] & L_n[f] \\ \phi_0(x) & \cdots & \phi_n(x) & 0 \end{vmatrix}, \quad (n = 0, 1, 2, \cdots).$$

For, operate on the right-hand member of (11) with L_i ($0 \leq i \leq n$), and subtract the resulting last row from the row with index i , using the property $L_i[0] = 0$. Then expand in terms of the elements in this row of index i . There results $L_i[f]$, ($i = 0, 1, \cdots, n$). That is, the right side of (11) is a linear combination of ϕ_0, \cdots, ϕ_n satisfying the conditions (2). But system (2) has a unique solution $s_n(x)$; whence (11) follows.

Let us now return to the definition of a best approximation method. If we start with a set of linear operators M_n , which assign functions to functions: $M_n[u(x)] = u_n(x)$, then by choosing a sequence of numbers $\{x_n\}$, we get a method \mathcal{M} by setting $L_n[u(x)] = \{M_n[u(x)]\}_{x=x_n}$. In particular, we may have $x_n \equiv a$.

An interesting subclass is that where the operators M_n are obtained by iteration from a single one:

$$M_0 \equiv I \equiv \text{identity}, \quad M_1 = M, \quad M_2 = M(M) = M^2, \cdots, \quad M_n = M^n, \cdots;$$

and where we choose $x_n \equiv a$ (which we may take as 0). The case of Widder finds itself in this class, with $M[u(x)] \equiv du(x)/dx$.

For any method \mathcal{M} which is non-singular relative to the set $\{\phi_n(x) = x^n\}$, there will exist a unique basic set of polynomials $\Phi_n(x) = \sum_{i=0}^n \lambda_{ni} x^i$. If $M[u] = du(x)/dx$, then $\Phi_n(x) = x^n/n!$. Again, if $M[u] = u(x+1) - u(x)$, then $\{\Phi_n(x)\}$ is the set of Newton polynomials:

$$\Phi_0(x) = 1, \quad \Phi_1(x) = x, \quad \Phi_n(x) = \frac{x(x-1) \cdots (x-n+1)}{n!}.$$

And in general, by means of these "methods" we can define large classes of sets of polynomials.

In particular, consider the class of orthogonal Tchebycheff polynomial sets. Given a function $p(x)$, a Tchebycheff set $\{T_n(x)\}$ is defined by the relations

$$\int_a^b p(t) T_m(t) T_n(t) dt = \begin{cases} 0, & (m \neq n), \\ 1, & (m = n). \end{cases}$$

These are equivalent (except for an undetermined multiplier in $T_n(x)$) to

$$\int_a^b p(t) t^i T_n(t) dt = 0, \quad (i = 0, 1, \dots, n-1).$$

If we now define

$$M_i[u(x)] = \int_a^b p(t) (t-x)^i u(t) dt, \quad (i = 0, 1, \dots),$$

then

$$L_i[T_n(x)] = \{M_i[T_n(x)]\}_{x=0} = 0, \quad (i = 0, 1, \dots, n-1),$$

and, by properly normalising $T_n(x)$, $L_n[T_n(x)] = 1$; so that $T_n(x) = \Phi_n(x)$. We thus see that *all orthogonal Tchebycheff polynomial sets can be defined by our methods of best approximation.*

More generally, the same is true for least square functions: Given the linearly independent set of functions $\{\phi_n(x)\}$, the set $\{T_n(x)\}$ is to be defined by

$$\int_a^b [f(t) - T_n(t)]^2 dt = \text{minimum},$$

where $T_n(t)$ ranges over all "polynomials" $T_n(x) = c_{n0}\phi_0(x) + \dots + c_{nn}\phi_n(x)$. By forming suitable linear combinations $\Phi_n(x) = b_{n0}\phi_0(x) + \dots + b_{nn}\phi_n(x)$, $b_{nn} \neq 0$, we can make $\{\Phi_n\}$ a normal orthogonal set:

$$\int_a^b \Phi_m(t) \Phi_n(t) dt = \begin{cases} 0, & (m \neq n), \\ 1, & (m = n). \end{cases}$$

- Furthermore, a method \mathcal{M} can be found for which $\{\Phi_n\}$ is the basic set. For, we have only to define

$$L_i[u(x)] = \int_a^b \Phi_i(t) u(t) dt;$$

then $L_i[\Phi_n(x)] = \begin{cases} 0, & (i = 0, 1, \dots, n-1), \\ 1, & (i = n), \end{cases}$

so that $\{\Phi_n\}$ is a basic set. Now we can express $T_n(x)$ as a linear combination

in the Φ_n 's: $T_n(x) = f_{n0}\Phi_0 + \cdots + f_{nn}\Phi_n$. From the minimum property we find that

$$f_{ni} = f_i = \int_a^b f(t) \Phi_i(t) dt.$$

Again, $L_i[T_n] = f_{ni}L_i[\Phi_i] = f_i$,

and $L_i[f] = f_i$;

hence $L_i[T_n] = L_i[f]$, $(i = 0, 1, \cdots, n)$;

i. e., equations (2) are satisfied, so that the minimizing set $\{T_n(x)\}$ is identical with the best approximating set $\{s_n(x)\}$.

As a final example, consider the following class of methods: Let

$$J(t) \sim a_1 t + a_2 t^2 + \cdots, \quad (a_1 \neq 0)$$

be a formal power series, generating the operator

$$J[u(x)] = a_1 u'(x) + a_2 u''(x) + \cdots.$$

Now define

$$M_0[u(x)] \equiv u(x), M_1[u(x)] \equiv J[u(x)], \cdots, M_n[u(x)] \equiv J^n[u(x)], \cdots,$$

giving the method \mathcal{M} :

$$L_n[u(x)] = \{M_n[u(x)]\}_{x=0}.$$

We have already pointed out the cases $J(t) = t$, giving $J[u] = du/dx$, and $J(t) = e^t - 1$, giving $J[u] = u(x+1) - u(x)$. For the general J there will exist a set of best approximating polynomials $\{\Phi_n(x)\}$, which is of considerable interest in the study of functional equations based on the operator $J[u(x)]$. This aspect of these polynomial sets will not concern us here, but we wish to point out an interesting recurrence relation among the polynomials of the set $\{\Phi_n(x)\}$. It is this:¹⁰

$$J[\Phi_n(x)] = \Phi_{n-1}(x), \quad (n = 1, 2, \cdots).$$

Let us turn once again to the Widder case. We can write

¹⁰ The following two particular cases are well-known:

$$\frac{d}{dx} \left[\frac{x^n}{n!} \right] = \frac{x^{n-1}}{(n-1)!};$$

$$\Delta \left[\frac{x(x-1) \cdots (x-n+1)}{n!} \right] = \frac{x(x-1) \cdots (x-n+2)}{(n-1)!}.$$

$$\Omega_n(x) = \frac{x^n}{n!} [1 + h_n(x)] = \frac{x^n}{n!} + c_{n,n+1} \frac{x^{n+1}}{(n+1)!} + c_{n,n+2} \frac{x^{n+2}}{(n+2)!} + \cdots;$$

or, on setting $\Phi_n(x) = x^n/n!$:

$$\Omega_n(x) = \Phi_n(x) + c_{n,n+1}\Phi_{n+1}(x) + c_{n,n+2}\Phi_{n+2}(x) + \cdots.$$

That is, $\{\Phi_n\}$ is a basic set for $\{\phi_n(x) = x^n\}$, and $\{\Omega_n\}$ is a basic set for some sequence, say $\{\omega_n(x)\}$; and $\Omega_n(x)$ has the above expression in terms of the set $\{\Phi_n\}$. This is a fairly general phenomenon, as the following theorem will show.

Let \mathcal{M} be a given method, and $\{\phi_n(x)\}$ a sequence relative to which \mathcal{M} is non-singular. Then, as we have seen, there exists a unique basic set $\{\Phi_n\}$.

THEOREM 2. Let $\{\phi_n(x)\}$, $\{\omega_n(x)\}$ be two sequences for which \mathcal{M} is non-singular, and let $\{\Phi_n(x)\}$, $\{\Omega_n(x)\}$ be their basic sets. If

(i) each $\omega_n(x)$ has a convergent Φ_n -expansion in a region \mathcal{R} :

$$(12) \quad \omega_n(x) = \sum_{i=0}^{\infty} a_{ni} \Phi_i(x);$$

(ii) the operators L_0, L_1, \cdots (which define \mathcal{M}) are term-wise applicable to the above expansions:

$$(13) \quad L_m[\omega_n(x)] = \sum_{i=0}^{\infty} a_{ni} L_m[\Phi_i(x)];$$

then the set $\{\Omega_n(x)\}$ has the form (convergent in \mathcal{R})

$$(14) \quad \Omega_n(x) = \Phi_n(x) + c_{n,n+1}\Phi_{n+1}(x) + c_{n,n+2}\Phi_{n+2}(x) + \cdots.$$

We observe first that since Ω_n is a linear combination of $\omega_0, \cdots, \omega_n$, it possesses a Φ_n -expansion convergent in \mathcal{R} : $\Omega_n(x) = \sum_{i=0}^{\infty} c_{ni} \Phi_i(x)$. Again, condition (ii)

allows term-wise operation by L_m on this series: $L_m[\Omega_n] = \sum_{i=0}^{\infty} c_{ni} L_m[\Phi_i]$.

Now $\{\Omega_n\}$, $\{\Phi_n\}$ are basic sets, so (5) holds for them. Taking $m = 0, 1, \cdots, n$, this yields the relations

$$c_{n0}L_m[\Phi_0] + c_{n1}L_m[\Phi_1] + \cdots + c_{nm}L_m[\Phi_m] = \begin{cases} 0, & (m = 0, 1, \cdots, n-1), \\ 1, & (m = n); \end{cases}$$

which, on recalling that $L_m[\Phi_m] = 1$, gives the following values for c_{ni} : $c_{n0} = c_{n1} = \cdots = c_{n,n-1} = 0$, $c_{nn} = 1$. Hence (14) holds.

2. *Convergence in the Widder case.*¹¹ We are here concerned with the convergence properties of Ω_n -series, where

¹¹ We have already remarked that in this section we shall lighten one of Widder's conditions. We add that the method used here is more direct than that of Widder.

$$(15) \quad \Omega_n(x) = (x^n/n!) [1 + h_n(x)] = (x^n/n!) \Theta_n(x), \quad h_n(0) = 0,$$

with $h_n(x)$ analytic in $|x| \leq R$.

THEOREM 3. Suppose constants c, N, β_n exist such that

$$(i) \quad 0 < c \leq |\Theta_n(x)| \leq \beta_n,$$

uniformly in $|x| \leq R$ for all $n > N$, with ¹²

$$(ii) \quad \limsup_{n \rightarrow \infty} \beta_n^{1/n} \leq 1.$$

If the series

$$(16) \quad f(x) = \sum_{n=0}^{\infty} f_n \Omega_n(x)$$

converges for a single point $x = \xi$ in $|x| \leq R$, it converges uniformly and absolutely in every closed region lying ¹³ in $|x| < |\xi|$, thus representing an analytic function in $|x| < |\xi|$.

For:

$$\sum_0^{\infty} |f_n \Omega_n(x)| = \sum_0^{\infty} |f_n \Omega_n(\xi)| |\Omega_n(x)/\Omega_n(\xi)| \leq A(x) + (M/c) \sum_{n=N}^{\infty} \beta_n |x/\xi|^n,$$

where $A(x)$ is the sum of the absolute values of the first N terms, and M is a bound (which exists) of $|f_n \Omega_n(\xi)|$. $\Omega_n(\xi)$ can vanish only if $\xi = 0$ or $\Theta_n(\xi) = 0$. Now the theorem is vacuously true if $\xi = 0$; and $\Theta_n(\xi) \neq 0$ by virtue of (i). Hence we can assume that $\Omega_n(\xi) \neq 0$; and the indicated division is possible. Since ¹⁴ $\limsup \beta_n^{1/n} \leq 1$, the last series converges uniformly and absolutely in every closed region in $|x| < |\xi|$, and this is then true of the original series.

It is seen that condition (i), although applying throughout $|x| \leq R$, is used only at the point $x = \xi$. Now it may happen that for some points in $|x| \leq R$ a number c (depending on the point) exists, and for other points it does not. This suggests strengthening Theorem 3 as follows:

THEOREM 3'. Let Σ be the set of those points $x = \xi$ in $|x| \leq R$ for which c, N, β_n exist (as functions of ξ) such that

A class of Ω_n -series (that arose in the study of some linear differential equations) is the \mathcal{D}_n -series of Transactions, *American Mathematical Society*, vol. 35 (1933), pp. 184-214.

¹² Incidentally, (ii) combined with $c \leq \beta_n$ gives $\limsup \beta_n^{1/n} = 1$.

¹³ It follows that if the region of convergence does not go outside of $|x| \leq R$, then it is a circular region.

¹⁴ If condition (ii) is replaced by (ii') $\limsup \beta_n^{1/n} \leq K$, then the region for which convergence can be asserted is $|x| < |\xi|/K$.

$$(i) \quad 0 < c(\xi) \leq |\Theta_n(\xi)| \leq \beta_n(\xi),$$

with

$$(ii) \quad \limsup_{n \rightarrow \infty} [\beta_n(\xi)]^{1/n} \leq 1.$$

If series (16) converges for a single point $x = \xi$ in Σ , it converges uniformly and absolutely in every closed region lying in $|x| < |\xi|$, to an analytic function.

The proof of Theorem 3 applies to 3'.

LEMMA 2. A function cannot have two $\Omega_n(x)$ -expansions uniformly convergent in an open region \mathcal{R} containing the point $x = 0$.

For on subtracting we should have $0 = \Sigma a_n \Omega_n(x)$, uniformly convergent in \mathcal{R} . By successive term-wise differentiations (which are permissible) at $x = 0$, we find that $(a_n = 0, n = 0, 1, \dots)$.

Let us try to develop the function $1/(t-x)$ (t a parameter) in an Ω_n -series. Assume that

$$(17) \quad 1/(t-x) = \sum_{n=0}^{\infty} L_n(t) \Omega_n(x).$$

By (formal) term-wise differentiation and setting $x = 0$, we get

$$(18) \quad \begin{aligned} 0!/t &= L_0(t) \\ 1!/t^2 &= L_0(t)h'_0(0) + L_1(t) \\ n!/t^{n+1} &= L_0(t)h_0^{(n)}(0) + L_1(t) \binom{n}{1} h_1^{(n-1)}(0) \\ &\quad + \dots + L_{n-1}(t) \binom{n}{n-1} h'_{n-1}(0) + L_n(t) \end{aligned}$$

thus determining the functions $\{L_n(t)\}$. $L_n(t)$ is, in fact, a polynomial in $1/t$ of degree $n+1$.

Define λ_n as the maximum of $|h_n(x)|$ in $|x| \leq R$:

$$(19) \quad |h_n(x)| \leq \lambda_n, \quad |x| \leq R.$$

Then

$$(20) \quad |h_n^{(m)}(0)| \leq m! \lambda_n / R^m.$$

Let r be any positive number, and let $\rho = \min(r, R)$. A simple application of (20) to (18) yields the inequalities

$$\begin{aligned} |L_0(t)| &\leq 0!/\rho; & |L_1(t)| &\leq (1!/\rho^2)(1 + \lambda_0); \\ |L_2(t)| &\leq (2!/\rho^3)(1 + \lambda_0)(1 + \lambda_1), \end{aligned}$$

uniform in $|t| \geq r$; and a straightforward induction proof gives

LEMMA 3. *For all n , and uniformly in $|t| \geq r$, where r is any positive number and $\rho = \min(r, R)$,*

$$(21) \quad |L_n(t)| \leq (n!/\rho^{n+1})(1 + \lambda_0)(1 + \lambda_1) \cdots (1 + \lambda_{n-1}).$$

Then, for $|t| \geq r$, $|x| \leq R$,

$$\sum_0^\infty |L_n(t)\Omega_n(x)| = \sum_0^\infty |L_n(t)| \frac{x^n}{n!} [1 + h_n(x)] \ll \frac{1}{\rho} \sum_0^\infty \left\{ \prod_{i=0}^n (1 + \lambda_i) \right\} \left| \frac{x}{\rho} \right|^n.$$

If u_n is the n -th term of the series on the right, $u_{n+1}/u_n = (1 + \lambda_{n+1})|x/\rho|$.

THEOREM 4. *Consider the series $\sum_0^\infty L_n(t)\Omega_n(x)$, where $\{L_n(t)\}$ is given by (18), and $|h_n(x)| \leq \lambda_n$, $|x| \leq R$. If $\limsup \lambda_n = K$ (finite), the series converges uniformly and absolutely in $|x| \leq l$, $|t| \geq r$, where r is any positive number, $\rho = \min(r, R)$, and l is any positive number less than $\rho/(K + 1)$; and the series represents, in this region, the function $1/(t - x)$.*

The convergence properties stated follow from the preceding relations. Let $H(x, t)$ be the sum of the series; it is analytic in x and t in $|x| \leq l$, $|t| \geq r$. Term-wise differentiation in x (which is permissible) with $x = 0$ gives

$$\left\{ \frac{\partial^n H(x, t)}{\partial x^n} \right\}_{x=0} = n!/t^{n+1} = \left\{ \frac{\partial^n [1/(t - x)]}{\partial x^n} \right\}_{x=0}, \quad (n = 0, 1, \cdots);$$

hence H coincides with $1/(t - x)$.

Especially interesting is the case $K = 0$, in which case $\limsup \lambda_n = \lim \lambda_n = 0$:

THEOREM 5. *If $\lim \lambda_n = 0$, then series (17) is valid, converging uniformly and absolutely in $|x| \leq l$, $|t| \geq r$, where $r > 0$ is arbitrary, $\rho = \min(r, R)$, and l is any positive number less than ρ .*

THEOREM 6. *If¹⁵ $\lim \lambda_n = 0$, every function $f(x)$, analytic about $x = 0$, possesses an Ω_n -expansion. If the distance from $x = 0$ to the nearest singularity of $f(x)$ is a , and if $\sigma = \min(a, R)$, this expansion is uniformly and absolutely convergent in $|x| \leq \tau$ where τ is any positive number $< \sigma$, and the coefficients of the expansion are given by*

$$(22) \quad f(x) = \sum_0^\infty f_n \Omega_n(x), \quad f_n = (1/2\pi i) \int_C f(t) L_n(t) dt,$$

¹⁵ If $\limsup \lambda_n = K$ (finite), every $f(x)$, analytic about $x = 0$, has an Ω_n -expansion, but with reduced radius of convergence.

C being any contour around $t=0$ and within $|t| < \sigma$. Moreover, $f(x)$ has only one Ω_n -expansion.

The convergence property follows at once on multiplying (17) through by $f(t)$ and integrating around C , using the Cauchy integral formula. All that remains is the uniqueness proof. From $\lim \lambda_n = 0$ follows the existence of β_n satisfying the conditions of Theorem 3. Hence if $f(x)$ possesses an expansion different from (22) and convergent in at least one point $x = \xi \neq 0$, it converges uniformly in an open region containing $x = 0$. Lemma 2 now applies, to give us a contradiction; hence there is uniqueness.

COROLLARY. The sets $\{\Omega_n\}$, $\{L_n\}$ are normal-orthogonal on C , any contour in $|t| < R$, surrounding $t=0$:

$$(1/2\pi i) \int_C \Omega_m(t) L_n(t) dt = \begin{cases} 0, & (m \neq n), \\ 1, & (m = n). \end{cases}$$

For, $\Omega_m(x)$ is analytic in $|x| \leq R$ and therefore possesses a unique Ω_n -expansion, the coefficient of $\Omega_n(x)$ being the above integral. Hence normality and orthogonality hold.

Theorem 6 can be given a different form:

THEOREM 7. Let the condition $\lim \lambda_n = 0$ be replaced by the following condition: $\Theta_n(x) = 1 + h_n(x)$ converges uniformly in $|x| \leq R$ to a function $M(x)$ which is nowhere zero in $|x| \leq R$. Then the conclusion of Theorem 6 is valid with the modification that f_n is now given by

$$(23) \quad f_n = (1/2\pi i) \int_C [f(t)/M(t)] L_n^*(t) dt,$$

where $\{L_n^*(t)\}$ is defined by (18) with $h_n(x)$ replaced by $g_n(x)$, the latter defined by

$$(24) \quad 1 + h_n(x) = M(x)[1 + g_n(x)].$$

- Clearly, $g_n(x) \rightarrow 0$ uniformly, $|x| \leq R$. The series $f(x) = \sum_0^\infty f_n \Omega_n(x)$ is identical with the series $f(x)/M(x) = \sum_0^\infty f_n(x^n/n!) [1 + g_n(x)]$, and since $\{\lambda_n\}$ exists such that $|g_n(x)| \leq \lambda_n$, $\lim \lambda_n = 0$, the second series expansion is valid by Theorem 6. The theorem now follows from the fact that the class of functions $\{f(x)\}$ analytic about $x=0$ coincides with the class of functions $\{f(x)/M(x)\}$.

3. *Extension of the Birkhoff theory.* The method of this section is

adapted from the work of Birkhoff, as was mentioned.¹⁶ We start with a set of functions $\{u_n(x)\}$, with given convergence properties, and seek to determine the convergence properties of a second set of functions $\{v_n(x)\}$, related to $\{u_n(x)\}$ only quantitatively. We shall introduce certain assumptions labelled Condition A, B, C; and it is to be understood that once a Condition has been stated, it is to hold from then on to the end of the section.

Consider a sequence of functions $\{u_n(x)\}$ satisfying Condition A:

(i) $u_n(x)$ is analytic in the interior \mathfrak{A} of a rectifiable, simple closed curve C , and is continuous in $\mathfrak{A} + C$.

(ii) The function $1/(t-x)$, t a parameter, possesses a u_n -expansion

$$(24) \quad 1/(t-x) = \sum_{n=0}^{\infty} L_n(t)u_n(x),$$

which is uniformly convergent in x and t for t on C and x on any closed point set in \mathfrak{A} ; and the functions $\{L_n(t)\}$ are continuous on C .¹⁷

COROLLARY. Every function $f(x)$ that is analytic in \mathfrak{A} and continuous in $\mathfrak{A} + C$, has a $u_n(x)$ -expansion, uniformly convergent on every closed point set in \mathfrak{A} :

$$(25) \quad f(x) = \sum_{n=0}^{\infty} f_n u_n(x), \quad (26) \quad f_n = (1/2\pi i) \int_C f(t) L_n(t) dt.$$

COROLLARY. If $\sum |L_n(t)u_n(x)|$ converges uniformly, t on C and x on any closed set in \mathfrak{A} , then series (25) converges absolutely, x in \mathfrak{A} .

Now consider the set $\{v_n(x)\}$, which is to be "close" to the set $\{u_n(x)\}$ in a sense to be defined. We assume that $v_n(x)$ is analytic in \mathfrak{A} and continuous in $\mathfrak{A} + C$. Suppose we have the expansion

$$(27) \quad f(x) = \sum_0^{\infty} \phi_n v_n(x).$$

¹⁶ A number of papers have been written on subjects related to the work of Widder and of Birkhoff. References are to be found in Widder's paper and also in: Walsh, *Transactions of the American Mathematical Society*, vol. 31 (1929), pp. 53-57. In this section we do not emphasize generality of statement in our theorems. Rather, we aim to secure a comprehensive body of theorems that are symmetric (i. e., interchangeable) in the two sets of functions $\{u_n\}$, $\{v_n\}$; and that can be utilized in treating convergence of series of best approximating "polynomials."

¹⁷ In Birkhoff's case, $u_n(x) = x^n$, so that Condition A holds when C is any circle with center at $x = 0$.

In analogy with (25, 26), we are led to consider the possibility¹⁸ of defining the coefficients ϕ_n by

$$(28) \quad \phi_n = (1/2\pi i) \int_C g(t) L_n(t) dt,$$

$g(t)$ to be determined.

We see from (28) and (25, 26) that

$$(29) \quad g(x) = \sum_{n=0}^{\infty} \phi_n u_n(x).$$

Substitution of (28) into (27) yields

$$(30) \quad f(x) = (1/2\pi i) \int_C \left\{ \sum_{n=0}^{\infty} v_n(x) L_n(t) \right\} g(t) dt.$$

This integral equation is not well-adapted to determine $g(t)$. We can obtain an equation of the second kind by subtracting from (30) the relation

$$(31) \quad g(x) = \frac{1}{2\pi i} \int_C \frac{g(t)}{t-x} dt = \frac{1}{2\pi i} \int_C \left\{ \sum_{n=0}^{\infty} u_n(x) L_n(t) \right\} g(t) dt.$$

In fact, we then have

$$(32) \quad f(x) = g(x) + (1/2\pi i) \int_C K(x, t) g(t) dt$$

where

$$(33) \quad K(x, t) = \sum_{n=0}^{\infty} [v_n(x) - u_n(x)] L_n(t).$$

We now assume

Condition B. *Series (33) converges uniformly for x in $\mathfrak{A} + C$ and t on C , and*

$$|K(x, t)| < 2\pi/l$$

for x and t on C , where l = length of C .

It is Condition B that is the test of $\{v_n(x)\}$ being "close" to $\{u_n(x)\}$.

COROLLARY. *$K(x, t)$ is analytic in x in the region \mathfrak{A} for each t on C , and is continuous in x and t for t on C and x in $\mathfrak{A} + C$.*

We observe that the formal work from (27) to (33) is valid if we work backwards; i. e., given $g(x)$, assumed to be analytic in \mathfrak{A} and continuous in $\mathfrak{A} + C$; then (28), (29), (31) hold, and (29) is uniformly convergent for x on any closed set in \mathfrak{A} . If now $f(x)$ is defined by (32), then $f(x)$ is seen

¹⁸ Our argument is purely formal until our conclusions are stated and proved.

to be analytic in \mathfrak{A} and continuous in $\mathfrak{A} + C$; and by combining (31) and (32), then (30) holds, the series within the brace being uniformly convergent in x and t for t on C and x on any closed set in \mathfrak{A} . (30) may then be integrated term-wise, yielding series (27), which is uniformly convergent for x on any closed set in \mathfrak{A} .

Our aim, however, is to start with $f(x)$ and determine $g(x)$. In (32) let x be chosen on C . As x and t traverse C they are functions of the arc length (measured from some point on C):

$$x = \Theta(s), \quad t = \Theta(\sigma).$$

Our hypothesis on C assures us of the existence of $d\Theta(\sigma)/d\sigma$ almost everywhere.¹⁹ Point x being on C , let us set

$$F(s) = f(\Theta(s)), \quad G(s) = g(\Theta(s)), \quad K^*(s, \sigma) = (1/2\pi i) K(\Theta(s), \Theta(\sigma)) \Theta'(\sigma).$$

Equation (32) then reduces to the equivalent form

$$(32') \quad F(s) = G(s) + \int_0^l K^*(s, \sigma) G(\sigma) d\sigma.$$

$F(s)$ is continuous; and so is $K^*(s, \sigma)$ except on a set of measure zero (due to the possible non-existence of $\Theta'(\sigma)$), where it can be defined so as to be bounded for s, σ in $0 \leq s, \sigma \leq l$. The Fredholm theory can be applied. Because of the second part²⁰ of Condition B, the Neumann series for a solution of (32') converges uniformly, so that $\lambda = 1$ is not a characteristic number. Hence (32') has a unique solution $G(s)$; and it is continuous, $0 \leq s \leq l$.

This continuous function $G(s)$ defines a continuous function $g(x)$ (x on C) satisfying (32); and $g(x)$ is a unique solution of (32). We now extend the definition of $g(x)$ to \mathfrak{A} by means of (32), where x is now in \mathfrak{A} . $g(x)$ is seen to be analytic in \mathfrak{A} . Suppose x , in \mathfrak{A} , approaches a point α of C .

The functions $f(x)$, $(1/2\pi i) \int_C K(x, t) g(t) dt$ being continuous in $\mathfrak{A} + C$, it follows from (32) that $g(x) \rightarrow g(\alpha)$, where $g(\alpha)$ is the value, at $x = \alpha$, of the unique solution of (32) for x on C . This gives us

¹⁹ And at such points where $\Theta'(\sigma)$ fails to exist, the difference quotient is nevertheless bounded: $|\Delta\Theta/\Delta\sigma| \leq 1$. Where Θ' does exist, it has the value $\Theta'(\sigma) = e^{i\theta}$ where θ is the angle which the tangent to C (at the point σ) makes with the real axis.

²⁰ Cf. Whittaker and Watson, *Modern Analysis*, 4th ed., pp. 221-222. A remark of Birkhoff (*loc. cit.*) is apropos here: It is not necessary that $|K(x, t)|$ be less than $2\pi/l$. All the work of this section will hold if we merely assume that in the integral equations (32) and (40), $\lambda = 1$ is not a characteristic number.

LEMMA 4. *To every function $f(x)$, analytic in \mathfrak{A} and continuous in $\mathfrak{A} + C$, there corresponds a unique solution $g(x)$ of (32); and $g(x)$ is also analytic in \mathfrak{A} and continuous in $\mathfrak{A} + C$.*

Having obtained the function $g(x)$, the observation made after Condition B, on going from $g(x)$ to $f(x)$, enables us to state

THEOREM 8. *Every function $f(x)$, analytic in \mathfrak{A} and continuous in $\mathfrak{A} + C$, has a $v_n(x)$ -expansion*

$$(27) \quad f(x) = \sum_{n=0}^{\infty} \phi_n v_n(x), \quad (28) \quad \phi_n = (1/2\pi i) \int_C g(t) L_n(t) dt,$$

uniformly convergent on every closed set in \mathfrak{A} . In (28), $g(t)$ is the function of Lemma 4.

COROLLARY. *If series $\sum |u_n(x) L_n(t)|$ converges uniformly for x in \mathfrak{A} and t on C , and series $\sum |v_n(x) - u_n(x)| L_n(t)|$ converges uniformly for x in $\mathfrak{A} + C$ and t on C , then series (27) (with ϕ_n given by (28)) converges absolutely for all x in \mathfrak{A} .*

For: We have $f(x) = \sum \phi_n v_n(x)$, $g(x) = \sum \phi_n u_n(x)$, so that

$$f - g = \sum (1/2\pi i) \int_C g(t) L_n(t) [v_n(x) - u_n(x)] dt.$$

By hypothesis this series converges absolutely. Again, from the second Corollary to Condition A, $g = \sum (1/2\pi i) \int_C g(t) L_n(t) u_n(x) dt$ converges absolutely. Hence on adding, (27) converges absolutely, x in \mathfrak{A} .

The functions $\{L_n(t)\}$ are defined only on C , where they are continuous. We can extend their definition to \mathcal{E} , the region exterior to C :

DEFINITION. *For z in \mathcal{E} , $L_n(z)$ is defined to be*

$$(34) \quad L_n(z) = \frac{-1}{2\pi i} \int_C \frac{L_n(t)}{t - z} dt.$$

COROLLARY. $L_n(z)$ is analytic ²¹ in \mathcal{E} , and $L_n(\infty) = 0$.

THEOREM 9. *The series*

$$(35) \quad 1/(z - x) = \sum_{n=0}^{\infty} L_n(z) u_n(x),$$

²¹ There is no reason for supposing, without further hypotheses, that L_n , defined in $\mathcal{E} + C$ by (34) and by Condition A, is continuous in $\mathcal{E} + C$. Later, when we do have a further condition, this assertion can be made. (See Theorem 20.)

holds uniformly in x and z for x on any closed set in \mathfrak{A} and z on any "closed" set²² in \mathfrak{E} .

To show this, observe that for z in \mathfrak{E} , $1/(z-x)$ is analytic (in x) in \mathfrak{A} and continuous in $\mathfrak{A} + C$, so that (Corollary to Condition A) it has a uniformly convergent $u_n(x)$ -expansion:

$$\frac{1}{z-x} = \sum_0^\infty \phi_n(z) u_n(x), \quad \phi_n(z) = \frac{1}{2\pi i} \int_C \left(\frac{1}{z-t} \right) L_n(t) dt.$$

From (34) we see that $\phi_n(z) = L_n(z)$. Since $\sum_0^\infty u_n(x) L_n(t)$ converges uniformly, x on a closed set in \mathfrak{A} and t on C , therefore term-wise integration (after multiplication by $1/(z-t)$) is permissible, the resulting series being uniformly convergent for x and z in the regions stated.

THEOREM 10. *If $f(z)$ is analytic in $\mathfrak{E} + C$, it has the $L_n(z)$ -expansion*

$$(36) \quad f(z) - f(\infty) = \sum_{n=0}^\infty \alpha_n L_n(z), \quad (37) \quad \alpha_n = \frac{1}{2\pi i} \int_C u_n(t) f(t) dt,$$

uniformly convergent for z on any "closed" set in \mathfrak{E} .

For: We can find a rectifiable simple closed curve J inside C such that $f(z)$ is analytic on J and exterior to J . Then, by (35),

$$f(z) - f(\infty) = \frac{-1}{2\pi i} \int_J \frac{f(x)}{x-z} dx = \sum_0^\infty \left\{ \frac{1}{2\pi i} \int_J u_n(x) f(x) dx \right\} L_n(z),$$

the series being uniformly convergent for z on a "closed" set in \mathfrak{E} . Now J can be chosen as close to C as we wish; whence it follows, from the continuity of $u_n(x)$ and $f(x)$ in the closed region consisting of J , C and the ring that they bound, that in the coefficient of $L_n(z)$ the curve of integration J may be replaced by C without altering values. That is, (37) holds.

Consider again equation (33). The series being uniformly convergent for x in $\mathfrak{A} + C$ and t on C , we may multiply through by $1/(t-z)$, z in \mathfrak{E} , and integrate term-wise:

$$\begin{aligned} \frac{-1}{2\pi i} \int_C \frac{K(x, t)}{t-z} dt &= \frac{-1}{2\pi i} \int_C \sum_0^\infty [v_n(x) - u_n(x)] \frac{L_n(t)}{t-z} dt \\ &= \sum_0^\infty [v_n(x) - u_n(x)] \left(\frac{-1}{2\pi i} \int_C \frac{L_n(t)}{t-z} dt \right), \end{aligned}$$

²² By a "closed" set in \mathfrak{E} we shall mean both the usual closed set and also any unbounded set in \mathfrak{E} (including $z = \infty$), the important feature being that the set is at a positive distance from C .

the resulting series being uniformly convergent for x in $\mathfrak{A} + C$, and z on any "closed" set in \mathfrak{E} . Now by (34), the parenthesis has the value $L_n(z)$. This enables us to extend the definition of K to \mathfrak{E} :

LEMMA 5. *The series*

$$(37) \quad K(x, z) = \sum_0^{\infty} [v_n(x) - u_n(x)] L_n(z)$$

converges uniformly in x and z for x in $\mathfrak{A} + C$ and z on any "closed" set in \mathfrak{E} , so that $K(x, z)$ is analytic in x and z for x in \mathfrak{A} and z in \mathfrak{E} . Moreover,

$$(38) \quad K(x, z) = \frac{-1}{2\pi i} \int_C \frac{K(x, t)}{t - z} dt,$$

where $K(x, t)$ is given by series (33); and $K(x, \infty) = 0$.

In the expansion (27), the coefficients ϕ_n are given in terms of $g(x)$. The question arises if we can express ϕ_n directly in terms of $f(x)$:

$$(39) \quad \phi_n = (1/2\pi i) \int_C f(t) M_n(t) dt,$$

where the functions $M_n(t)$ are to be determined. Since (32) holds for x in $\mathfrak{A} + C$, we may substitute for $f(t)$ in (39) its value as given by (32). On further using the relation $\phi_n = (1/2\pi i) \int_C g(t) L_n(t) dt$, this gives

$$(a) \quad (1/2\pi i) \int_C g(t) \{L_n(t) - M_n(t) - (1/2\pi i) \int_C K(w, t) M_n(w) dw\} dt = 0.$$

Now (a) is to hold for all g , and we want M_n to be independent of f (and therefore of g). This suggests that we set the brace equal to zero:

$$(40) \quad L_n(t) = M_n(t) + (1/2\pi i) \int_C K(w, t) M_n(w) dw.$$

This integral equation can be thrown into "real" form, as was (32). The resulting kernel is $K^{**}(s, \sigma) = (1/2\pi i) K(\Theta(\sigma), \Theta(s)) \Theta'(\sigma)$, so, as was the case with (32), (40) has a unique solution $M_n(t)$ (t on C), and $M_n(t)$ is continuous.

THEOREM 11. *In Theorem 8, the coefficients ϕ_n can also be expressed by (39), where $M_n(t)$ is the unique and continuous solution of (40), t on C .*

To see this, we observe first that (a) is satisfied. On using (28) and (32), (a) simplifies to

$$\begin{aligned}
 0 &= \phi_n - (1/2\pi i) \int_C M_n(t) [g(t) + (1/2\pi i) \int_C K(t, w) g(w) dw] dt \\
 &= \phi_n - (1/2\pi i) \int_C M_n(t) f(t) dt,
 \end{aligned}$$

which is (39).

By means of (40), with t replaced by z , we can extend the definition of M_n :

DEFINITION. For z in \mathcal{E} ,

$$(41) \quad M_n(z) = L_n(z) - (1/2\pi i) \int_C K(t, z) M_n(t) dt,$$

$M_n(t)$ being the unique and continuous solution of (40).

COROLLARY. $M_n(z)$ is analytic in z , with $M_n(\infty) = 0$; and for z in \mathcal{E} ,

$$(42) \quad M_n(z) = \frac{-1}{2\pi i} \int_C \frac{M_n(t)}{t-z} dt.$$

We need only establish (42). If we multiply (40) through by $[(-1)/2\pi i] \times [1/(t-z)]$ and integrate over C , and use (34) and (38), we get

$$L_n(z) = \frac{-1}{2\pi i} \int_C \frac{M_n(t)}{t-z} dt + \frac{1}{2\pi i} \int_C K(w, z) M_n(w) dw.$$

Comparison with (41) then yields (42).

THEOREM 12. The expansion

$$(43) \quad 1/(z-x) = \sum_0^{\infty} v_n(x) M_n(z)$$

is uniformly convergent in x and z for x on any closed set in \mathfrak{A} and z on any "closed" set in \mathcal{E} .

For: In (32), choose $f(x) = f(x, z) = 1/(z-x)$, z in \mathcal{E} . Then $g(x) = g(x, z)$ is defined by (32) to be analytic in x and z for x in \mathfrak{A} and z in \mathcal{E} ; and is continuous for x in $\mathfrak{A} + C$ and z in \mathcal{E} . If we multiply series (24) through by $(1/2\pi i) \cdot [1/(z-t)]$ and integrate term-wise around C , we observe that $g(x, z)$ has a $u_n(x)$ -expansion (cf. (29)) that is uniformly convergent in x and z , x on any closed set in \mathfrak{A} and z on any "closed" set in \mathcal{E} :

$$g(x, z) = \sum_{n=0}^{\infty} u_n(x) \phi_n(z), \quad \phi_n(z) = (1/2\pi i) \int_C L_n(t) g(t, z) dt.$$

If we now combine (31, 32, 33), where $g(x) = g(x, z)$ (so that $f(x) = f(x, z) = 1/(z-x)$), and recall that (31) converges uniformly in

both x and z , then (30) is seen to hold, also uniformly convergent in both x and z . Term-wise integration gives (27), again uniformly convergent in x and z : $1/(z-x) = \sum_0^\infty \phi_n(z)v_n(x)$. It remains to identify the coefficient of $v_n(x)$. This coefficient is the coefficient ϕ_n given by (27) and (28); and by Theorem 11 this coefficient has the value given by (39):

$$\phi_n(z) = \frac{1}{2\pi i} \int_C \frac{M_n(t)}{z-t} dt.$$

Comparison with (42) shows that $\phi_n(z) = M_n(z)$, and the theorem is established.

From this follows (cf. Theorem 10)

THEOREM 13. *If $f(z)$ is analytic in $\mathcal{E} + C$, it has the $M_n(z)$ -expansion*

$$(44) \quad f(z) - f(\infty) = \sum_{n=0}^\infty \beta_n M_n(z) dz; \quad (45) \quad \beta_n = (1/2\pi i) \int_C v_n(t) f(t) dt,$$

uniformly convergent for z on any "closed" set in \mathcal{E} .

There is apparent, by this time, a duality between the sets $\{u_n\}$, $\{L_n\}$ on the one hand and the sets $\{v_n\}$, $\{M_n\}$ on the other. We shall now examine to what extent their rôles can be interchanged. If $H(x, t)$ exists, having the relation to $\{v_n\}$ that $K(x, t)$ has to $\{u_n\}$, we should expect it to be given by the series

$$(45) \quad H(x, t) = \sum_{n=0}^\infty [u_n(x) - v_n(x)] M_n(t).$$

For the moment we shall put aside the problem of convergence of this series. If we substitute (40) into (33), we get the relation ²³

$$(46) \quad H(x, t) + K(x, t) + (1/2\pi i) \int_C H(x, w) K(w, t) dw = 0.$$

This is an integral equation for $H(x, t)$, with the same kernel $K(w, t)$ as in (40). Knowing the properties of $K(x, t)$, we can therefore state the following properties for $H(x, t)$:

LEMMA 6. *The function $H(x, t)$ defined by (46) is the only solution; it is continuous in x and t for x in $\mathcal{D} + C$ and t on C , and is analytic (in x) for x in \mathcal{D} and for each t on C ; and is continuous in x and t for x in $\mathcal{D} + C$ and t in \mathcal{E} , and is analytic (in x and t) for x in \mathcal{D} and t in \mathcal{E} .*

²³ This is the well-known equation for kernel and resolvent kernel in integral equation theory.

THEOREM 14. $K(x, t)$ is a resolvent kernel for the equation

$$(47) \quad g(x) = f(x) + (1/2\pi i) \int_C H(x, t) f(t) dt.$$

For, in the right-hand side of (47), let f have the value given by (32). On simplifying we obtain $g(x) + (1/2\pi i) \int_C g(t) A(x, t) dt$, where $A(x, t)$ is the left-hand member of (46); i. e., $A(x, t) \equiv 0$, so that the right-hand side of (47) does equal $g(x)$, and (47) is satisfied.

Since equation (47) has the solution (32) for every function $g(x)$ that is continuous on C , it follows from the Fredholm theory that (47) always has a unique solution. Hence

COROLLARY 1. For every $g(x)$, analytic in \mathfrak{D} and continuous in $\mathfrak{D} + C$, equation (47) has a unique solution $f(x)$. This solution is analytic in \mathfrak{D} and continuous in $\mathfrak{D} + C$.

It is also an immediate consequence of the uniqueness of solutions of both (32) and (47) that

COROLLARY 2. $H(x, t)$ is a resolvent kernel of (32); i. e., that the unique solution of (32) is furnished by (47).

If in (47) we set $f(x) \equiv K(x, w)$, we find on using (46) that $g(x) \equiv -H(x, w)$. Substituting these values of f and g into (32) then gives us the equation which is the twin of (46):

COROLLARY 3. The functions H and K satisfy the equation

$$(46') \quad H(x, t) + K(x, t) + (1/2\pi i) \int_C K(x, w) H(w, t) dw = 0,$$

valid for x in $\mathfrak{D} + C$ and t in $\mathfrak{E} + C$.

LEMMA 7. The unique solution $M_n(t)$ of equation (40) is given by

$$(48) \quad M_n(t) = L_n(t) + (1/2\pi i) \int_C H(w, t) L_n(w) dw.$$

This follows on using (46').

We can now establish the validity of (45).

THEOREM 15. $H(x, t)$ has the expansion (45), which converges uniformly for x in $\mathfrak{D} + C$ and t on C .

To show this, we first have series (33), uniformly convergent in the region stated. We therefore have (from (48)):

$$\begin{aligned} \sum_0^{\infty} [v_n(x) - u_n(x)] M_n(t) \\ &= \sum_0^{\infty} [v_n(x) - u_n(x)] \{L_n(t) + (1/2\pi i) \int_C H(w, t) L_n(w) dw\} \\ &= \sum_0^{\infty} [v_n(x) - u_n(x)] L_n(t) \\ &\quad + (1/2\pi i) \int_C H(w, t) \left\{ \sum_0^{\infty} [v_n(x) - u_n(x)] L_n(w) dw \right\}, \end{aligned}$$

the two series on the right converging uniformly in the region stated. Hence this same convergence property applies to the series on the left. There remains only to prove that this series has the value $-H(x, t)$. But this follows from (46') since the right-hand side is $K(x, t) + (1/2\pi i) \int_C H(w, t) K(x, w) dw$.

Theorem 8 can be dualized:

THEOREM 16. *If $g(x)$ is analytic in \mathfrak{A} and continuous in $\mathfrak{A} + C$, then we have*

$$(49) \quad g(x) = \sum_0^{\infty} \psi_n u_n(x), \quad (50) \quad \psi_n = (1/2\pi i) \int_C f(t) M_n(t) dt,$$

uniformly convergent on any closed in \mathfrak{A} . Here $f(x)$ is the solution of (47). For: In (47) replace $H(x, t)$ by its uniformly convergent expansion (45). There results the equation

$$g(x) = f(x) + \Sigma[(1/2\pi i) \int_C f(t) M_n(t) dt] [u_n(x) - v_n(x)],$$

the series being uniformly convergent on any closed set in \mathfrak{A} . But the series

$$f(x) = \Sigma[(1/2\pi i) \int_C f(t) M_n(t) dt] v_n(x)$$

has the same convergence property (Theorem 11); hence so has

$$\Sigma[(1/2\pi i) \int_C f(t) M_n(t) dt] u_n(x).$$

We thus have

$$g(x) = f(x) + \Sigma[(1/2\pi i) \int_C f(t) M_n(t) dt] u_n(x) - f(x),$$

from which (49) follows.

We have now an almost complete duality of $\{u_n\}$, $\{L_n\}$ and $\{v_n\}$, $\{M_n\}$. That it is not fully complete (at least so far as has been proved) is owing to

this lack; in Condition A we are not certain that (24) holds in the region stated, when u_n, L_n are replaced by v_n, M_n . What we do know, up to this point, is that (24) will hold (cf. (43)) if t is in \mathcal{E} , rather than on C ; nor does Theorem 8 permit t to be on C . However, we can fill in the gap:

THEOREM 17. *The expansion*

$$(50) \quad 1/(t-x) = \sum_{n=0}^{\infty} M_n(t) v_n(x)$$

is valid, and is uniformly convergent in x and t for x on any closed point set in \mathfrak{A} and t on C .

To show this, we begin with the expansion

$$(a) \quad 1/(t-x) = \sum_0^{\infty} L_n(t) u_n(x),$$

which has the convergence properties stated above. On multiplying through by $(1/2\pi i)H(t, w)$ we may integrate term-wise, the resulting series being likewise uniformly convergent:

$$(b) \quad \sum_0^{\infty} (1/2\pi i) \int_C H(w, t) \cdot u_n(x) L_n(w) dw.$$

Hence the series $\sum_0^{\infty} \{L_n(t) + (1/2\pi i) \int_C H(w, t) L_n(w) dw\} u_n(x)$ converges uniformly in x and t for x on any closed set in \mathfrak{A} and t on C . But the brace equals $M_n(t)$ (cf. (48)); hence the series $\sum_0^{\infty} M_n(t) u_n(x)$ converges uniformly in x and t in the region stated. Now (b) simplifies to

$$(1/2\pi i) \int_C H(w, t) \cdot [1/(w-x)] dw = H(x, t),$$

so that

$$\sum_0^{\infty} M_n(t) u_n(x) = 1/(t-x) + H(x, t) = 1/(t-x) + \sum_0^{\infty} [u_n(x) - v_n(x)] M_n(t).$$

The two series are uniformly convergent in x and t for x on any closed set in \mathfrak{A} and t on C . If then we subtract the first series from both members, we get (50), the series having the same convergence properties. This establishes the theorem.²⁴

²⁴ There is another point concerning duality: In Condition B we have $|K(x, t)| < 2\pi/l$. Now we do not know that H satisfies the same condition. In fact, if we write $\max |K(x, t)| = 2\pi\sigma/l$, $\sigma < 1$, then from (46) all we know is that $\max |H(x, t)|$

We may sum up, in part, as follows: If $\{u_n\}$ satisfies Conditions A and B, so ²⁵ does $\{v_n\}$.

The function $u_m(x)$, being analytic in \mathfrak{A} and continuous in $\mathfrak{A} + C$, has a u_n -expansion (cf. (25)):

$$u_m(x) = \sum_{n=0}^{\infty} f_{mn} u_n(x), \quad f_{mn} = (1/2\pi i) \int_C u_m(t) L_n(t) dt.$$

If a u_n -expansion is unique, then we have the biorthogonal property

$$(51) \quad (1/2\pi i) \int_C u_m(t) L_n(t) dt = \begin{cases} 0, & (m \neq n); \\ 1, & (m = n). \end{cases}$$

But Condition A does not insure uniqueness. For example, take $u_0(x) = 1$; $u_n(x) = (x^{n-1}/n!)(x - n)$, ($n > 0$). It is readily shown ²⁶ that if series $\sum c_n u_n(x)$ converges, the region of convergence is the interior of a circle, center the origin, reaching out to the nearest singularity of the function that is defined; and the convergence is uniform on any closed set within the circle of convergence. Moreover, every function, analytic about $x = 0$, has a u_n -expansion; and the functions $\{L_n\}$ can be defined by

$$L_n(t) = - \sum_{i=0}^{n-1} (i!/t^{i+1}), \quad (n > 0); \quad L_0(t) = -1/t.$$

Condition A is fulfilled on choosing C as any circle with center at $x = 0$.

But the function zero has the uniformly convergent expansion $0 = \sum_0^{\infty} u_n(x)$, so there fails to be uniqueness.

If there is to be uniqueness, it must appear in our assumptions. We accordingly add the uniqueness

Condition C. If zero has the expansion $0 = \sum_0^{\infty} a_n u_n(x)$, uniformly convergent on every closed set in \mathfrak{A} , then $a_n = 0$, ($n = 0, 1, \dots$).

From this follows that a function $f(x)$ cannot have two distinct u_n -expansions, each uniformly convergent on every closed set in \mathfrak{A} . Consequently we have

$< [\sigma/(1-\sigma)](2\pi/l)$. But the only use made of the condition $|K| < 2\pi/l$ is to secure uniqueness of the solutions of the integral equations with kernels $K(x, t)$ and $K(t, x)$. Hence we ought to establish this same uniqueness for the kernels $H(x, t)$ and $H(t, x)$. Now it is already known for $H(x, t)$ (cf. Theorem 14). That it also holds for $H(t, x)$ is easily shown.

²⁵ It being understood that in B the inequality $|K| < 2\pi/l$ is replaced by the assumption of uniqueness of the solutions for the kernels $K(x, t)$ and $K(t, x)$.

²⁶ Compare the proof of Theorem 3.

LEMMA 8. *The biorthogonality relations (51) hold.*

LEMMA 9. *The sets $\{v_n\}$, $\{M_n\}$ are biorthogonal:*

$$(52) \quad (1/2\pi i) \int_C v_m(t) M_n(t) dt = \begin{cases} 0, & (m \neq n); \\ 1, & (m = n). \end{cases}$$

Multiply equation (40) through by $u_m(t)$ and integrate over C :

$$\delta_{mn} = (1/2\pi i) \int_C M_n(t) u_m(t) dt + (1/2\pi i)^2 \int_C \int_C K(w, t) M_n(w) u_m(t) dw dt.$$

If we replace $K(w, t)$ by its uniformly convergent expansion (33), this reduces to

$$\delta_{mn} = (1/2\pi i) \int_C M_n(t) u_m(t) dt + (1/2\pi i) \int_C [v_m(w) - u_m(w)] M_n(w) dw,$$

and on cancelling the first and third terms (whose sum is zero), there remains relation (52).

THEOREM 18. *Equations (32) and (47) are satisfied by taking $f(x) = v_m(x)$, $g(x) = u_m(x)$.*

To see this, let $g(x) = u_m(x)$ in (32) and replace $K(x, t)$ by its expansion (33). Then, using (51), $f(x) = u_m(x) + [v_m(x) - u_m(x)] = v_m(x)$.

We come now to $\{v_n\}$ -uniqueness:

THEOREM 19. *If the v_n -expansion $\sum_{n=0}^{\infty} c_n v_n(x)$ converges uniformly in $\mathfrak{A} + C$ (the sum function $f(x)$ being therefore analytic in \mathfrak{A} and continuous in $\mathfrak{A} + C$), then necessarily $c_n = (1/2\pi i) \int_C f(t) M_n(t) dt$.*

Multiply the series through by $M_s(x)$ and integrate over C :

$$(1/2\pi i) \int_C f(t) M_s(t) dt = \sum_{n=0}^{\infty} c_n \cdot (1/2\pi i) \int_C v_n(t) M_s(t) dt.$$

By biorthogonality, the series on the right reduces to c_s , thus establishing the theorem.

Suppose we make a temporary translation of the complex variable x so as to insure that the origin is within C . This will not affect any of the results already obtained. Now it will be true that

$$\frac{1}{2\pi i} \int_C \frac{dt}{t^m(t-x)} = 0, \quad (m = 1, 2, \dots);$$

whence from the uniformly convergent expansion (24), we obtain

$$0 = \sum_{n=0}^{\infty} \left\{ \frac{1}{2\pi i} \int_C \frac{L_n(t)}{t^m} dt \right\} u_n(x), \quad (m = 1, 2, \dots),$$

uniformly convergent for x on any closed set in \mathfrak{A} . By the uniqueness property, all the coefficients must vanish:

$$(a) \quad \frac{1}{2\pi i} \int_C \frac{L_n(t)}{t^m} dt = 0, \quad \begin{cases} (n = 0, 1, \dots) \\ (m = 1, 2, \dots) \end{cases}$$

Now for a given n , condition (a) is necessary and sufficient²⁷ that there exist a function (which will of necessity be $L_n(z)$), analytic in \mathfrak{E} , and such that as $z \rightarrow t$ on C , $L_n(z) \rightarrow L_n(t)$.

It follows that $L_n(z)$ is continuous on $\mathfrak{E} + C$, and analytic in \mathfrak{E} . If we multiply (33) through by $(1/2\pi i) \cdot (1/t^m)$ and integrate around C , we see from (a) that

$$\frac{1}{2\pi i} \int_C \frac{K(x, t)}{t^m} dt = 0, \quad (m = 1, 2, \dots),$$

so that $K(x, z)$ also has the property that as z in \mathfrak{E} approaches a point t on C , then $K(x, z) \rightarrow K(x, t)$. That is, $K(x, z)$ is continuous for x in $\mathfrak{A} + C$ and z in $\mathfrak{E} + C$. Finally, from (40) we get

$$\frac{1}{2\pi i} \int_C \frac{M_n(t)}{t^m} dt = 0, \quad (m = 1, 2, \dots),$$

so that $M_n(z)$ is continuous in $\mathfrak{E} + C$; and therefore, also, $H(x, z)$. To sum up:²⁸

THEOREM 20. *The functions $L_n(z)$, $M_n(z)$, $K(x, z)$, $H(x, z)$ are continuous in x and z for x in $\mathfrak{A} + C$ and z in $\mathfrak{E} + C$; and their values for $z = t$ on C are respectively the known functions $L_n(t)$, $M_n(t)$, $K(x, t)$, $H(x, t)$.*

²⁷ If C is an analytic Jordan curve, this result holds if $\phi(t)$, the function given on the boundary (which in our case is $L_n(t)$) is continuous. (Walsh, *Transactions of the American Mathematical Society*, vol. 30 (1928), especially pp. 327 and 329.) If C is rectifiable, this same result holds if $\phi(t)$ is merely Lebesgue integrable (in which case the approach holds almost everywhere and must be non-tangential). (Priwaloff, *Comptes Rendus*, vol. 178 (1924), pp. 611-614.) In the Priwaloff case it is not clear (although probably true) that if $\phi(t)$ is continuous, then the approach holds everywhere on C . If this is not the case, we shall regard it as assumed that C satisfies the Walsh condition.

²⁸ If we now undo the translation that was made temporarily, none of these properties of continuity in $\mathfrak{E} + C$ will be altered.

In some of our theorems we had to insist on z being on any "closed" set in \mathcal{E} , because we had not this last theorem. It will be clear that we can now amend some of the theorems, as follows:

COROLLARY. *In Theorem 9, uniform convergence maintains for z in $\mathcal{E} + C$; Theorems 12 and 17 combine to give uniform convergence for z in $\mathcal{E} + C$; and Lemma 5 holds uniformly for z in $\mathcal{E} + C$.*

If we were to consider interchanging the rôles of $\{u_n\}$ and $\{L_n\}$, or of $\{v_n\}$ and $\{M_n\}$, we would define two functions $A(z, x)$, $B(z, x)$ by the series

$$(53) \quad \begin{aligned} A(z, x) &= \sum_{n=0}^{\infty} [M_n(z) - L_n(z)] u_n(x), \\ B(z, x) &= \sum_{n=0}^{\infty} [L_n(z) - M_n(z)] v_n(x). \end{aligned}$$

These functions are closely related to H and K . In fact we have the

COROLLARY. *The above series (53) converge uniformly in z and x for z on any "closed" set in \mathcal{E} and x in $\mathcal{D} + C$; and*

$$(54) \quad A(z, x) = H(x, z); \quad B(z, x) = K(x, z).$$

The convergence is immediate; and (54) follows from (33), (45) and (35), (43) (with reference to the preceding Corollary).

This is as far as we shall carry the theory of the $\{u_n\}$ -, $\{v_n\}$ -sets. We now point out how the results of the present section can be applied to the convergence question in methods of best approximation.

THEOREM 21. *Let $u_n(x) = \Phi_n(x)$, $v_n(x) = \Omega_n(x)$, ($n = 0, 1, \dots$) be two basic sets relative to a method \mathcal{M} of best approximation, and let $\Phi_n(x)$, $\Omega_n(x)$ be analytic in a region \mathcal{D} and continuous in $\mathcal{D} + C$, C being the boundary of \mathcal{D} . We further assume:*

(i) *Conditions A and B hold.*

(ii) *If ²⁰ $\{h_n(x)\}$ is any sequence of functions analytic in \mathcal{D} and continuous in $\mathcal{D} + C$, then the operators $\{L_i\}$ that define the method \mathcal{M} are term-wise applicable to every series $\sum c_n h_n(x)$ that converges uniformly on every closed point set in \mathcal{D} .*

²⁰ We actually make use of (ii) only for the sequences $h_n(x) = \Phi_n(x)$, $\Omega_n(x)$. The following observation is of interest: By the Corollary to Condition A, each $\Omega_n(x)$ has a $\Phi_n(x)$ -expansion uniformly convergent on any closed point set in \mathcal{D} . Therefore (ii) and Theorem 2)

$$(14) \quad \Omega_n(x) = \Phi_n(x) + c_{n,n+1} \Phi_{n+1}(x) + c_{n,n+2} \Phi_{n+2}(x) + \dots$$

From (14) it is seen that Condition B will certainly be fulfilled if the coefficients $c_{n,n+i}$ are chosen sufficiently small.

Under these conditions, if $f(x)$ is any function analytic in \mathfrak{A} and continuous in $\mathfrak{A} + C$, the approximating "polynomials" $s_n(x)$ of $f(x)$ (relative to the basic set $\{\Omega_n(x)\}$) converge uniformly to $f(x)$ on every closed point set in \mathfrak{A} .

For: By Theorem 8, $f(x)$ has an $\Omega_n(x)$ -expansion, uniformly convergent on any closed set in \mathfrak{A} : (a) $f(x) = \sum_0^\infty F_n \Omega_n(x)$. On the other hand, if $s_n(x)$ is the best "polynomial" of n -th order, then (using (6) with Φ_n replaced by Ω_n),

$$(b) \quad s_0(x) = f_0 \Omega_0(x), \quad s_n(x) - s_{n-1}(x) = f_n \Omega_n(x),$$

$$(c) \quad f_n = L_n[s_n(x) - s_{n-1}(x)].$$

Also ((2)),

$$(d) \quad L_i[s_i(x)] = L_i[f(x)].$$

The theorem will be established if we show that $F_n = f_n$, ($n = 0, 1, \dots$), since (cf. (8)) $s_n(x) = \sum_{i=0}^n f_i \Omega_i(x)$. By (ii) we may operate³⁰ with L_i on (a):

$$(e) \quad L_i[f] = F_0 L_i[\Omega_0] + F_1 L_i[\Omega_1] + \dots + F_i L_i[\Omega_i], \quad (i = 0, 1, \dots).$$

Taking $i = 0$: $L_0[F] = F_0 L_0[\Omega_0]$; $L_0[s_0] = F_0 L_0[\Omega_0]$; therefore $f_0 L_0[\Omega_0] = F_0 L_0[\Omega_0]$, and $f_0 = F_0$ since $L_0[\Omega_0] = 1$. Now assume that $F_r = f_r$, ($r = 0, 1, \dots, i-1$); we shall complete the induction for i : (e) reduces to

$$\begin{aligned} L_i[f] &= f_0 L_i[\Omega_0] + f_1 L_i[\Omega_1] + \dots + f_{i-1} L_i[\Omega_{i-1}] + F_i; \\ L_i[s_i] &= L_i[s_0] + L_i[s_1 - s_0] + \dots + L_i[s_{i-1} - s_{i-2}] + F_i; \\ L_i[s_i] &= L_i[s_{i-1}] + F_i; \quad L_i[s_i - s_{i-1}] = F_i; \end{aligned}$$

therefore $f_i = F_i$. This completes the proof.³¹

A comparison of (8) with (28) and (39) yields the

COROLLARY. The coefficients in the Ω_n -expansion for $f(x)$ are given (variously) by

$$(55) \quad f_n = \frac{1}{2\pi i} \int_C g(t) L_n(t) dt = \frac{1}{2\pi i} \int_C f(t) M_n(t) dt = L_n[s_n(x) - s_{n-1}(x)].$$

PENNSYLVANIA STATE COLLEGE.

³⁰ We also use (5): $L_i[\Omega_n] = 0$, $i < n$.

³¹ It is worth noting that in this theorem we do not assume the uniqueness Condition C. A possible choice of $v_n(x)$ is $v_n(x) \equiv u_n(x) = \Phi_n(x)$. Hence, if we omit the details that make Theorem 21 precise, the sense of the theorem is contained in the statement: *If $\{\Phi_n\}$, $\{\Omega_n\}$ are two basic sets "sufficiently close" to each other, they give essentially the same convergence properties to the respective best approximating "polynomials."*

GROUPS CONTAINING FIVE AND ONLY FIVE SQUARES.

By G. A. MILLER.

Let G represent a group such that the squares of its operators are five and only five distinct operators including the identity. When these five squares constitute a group it results that there are two and only two such groups which are not direct products. This is a special case of a general theorem relating to groups whose squares constitute a cyclic subgroup.¹ When these squares do not constitute a group there are three possible cases, as follows: Two of them are of order 4 and two of order 2, two are of order 3 and two of order 2, or four of them are of order 2. In each of these cases the identity constitutes the fifth square. This is always found among the squares since it is the square of itself.

When two of the squares are of order 4 then the squares generate the abelian group of order 8 and of type $(2, 1)$ which includes two operators of order 4 and one of order 2 which are non-squares. These 8 operators constitute an invariant subgroup of G which corresponds to an abelian quotient group of order 2^m and of type $(1, 1, 1, \dots)$. This invariant subgroup appears in an invariant abelian subgroup of order 16 and of type $(3, 1)$. Since G includes operators of order 4 whose common square is not equal to the square of the operators of order 4 which appear in this subgroup of order 16 the latter operators are not commutative with the former and they give rise to commutators of order 4 with respect to the operators of order 8 in the given subgroup of type $(3, 1)$. As these automorphisms are of order 2 these commutators are transformed into their inverses under G .

For the sake of brevity in the statements it will be assumed in what follows that G is not the direct product of a group containing five and only five operators which are squares and of an abelian group of order 2^m and of type $(1, 1, 1, \dots)$. The order of G cannot be less than 32 and when it is of this order its central is the four group contained in the given invariant abelian subgroup of type $(3, 1)$. There is one such G in which each of the operators of this invariant subgroup is transformed into its inverse and all of the remaining operators are of order 4 and have a common square which is distinct from the square of the operators of order 4 contained in the given

¹ G. A. Miller, *Proceedings of the National Academy of Sciences*, vol. 20 (1934), pp. 203-206.

subgroup of order 16. There is no such G in which each of the operators of this subgroup is transformed into its third power but there are two such G 's in which the commutator subgroup is the cyclic group of order 4 which is not generated by an operator of order 8 contained in G . In one of these two groups 8 of the additional operators are of order 2 while in the other all of the additional operators are of order 4 but have two distinct squares. Hence there results the following theorem: *Every group which has the property that it contains five and only five operators which are squares, including such an operator of order 4, involves at least one of the three groups of order 32 which have this property.*

To determine all the possible groups of order 64 which contain five and only five operators which are squares including one of order 4 it is therefore only necessary to extend each of the three groups noted in the preceding theorem by 32 additional operators. These include 16 operators which are commutative with an operator of order 4 which is a square and hence each possible set of 32 additional operators includes an operator of order 2 which is commutative with this operator of order 4. To the first of the three given groups of order 32 we can adjoin three such sets of 32 operators and thus obtain three G 's of order 64. In two of these the given added operator of order 2 has only two conjugates under G while in the third it has four such conjugates. To each of the other two given groups of order 32 we can adjoin only one such set of 32 operators. As all of these groups are distinct, there results the theorem that *there are five and only five groups of order 64 which separately have the property that they involve five and only five squares including such an operator of order 4.*

If such a G of order 128 exists it contains a subgroup of order 64 composed of all of its operators which are commutative with an operator t_1 of order 4 which is a square under G and all of whose operators of order 4 have a common square. Suppose first that this subgroup is abelian. An operator t_2 of order 4 in G whose square is different from t_1^2 is then commutative with at least 8 of the operators of this subgroup of order 64 and all of these operators besides the identity are of order 2. Hence G involves an operator of order 2 which is not contained in the subgroup generated by its squares but is commutative with all of its operators. It is therefore a direct product of an abelian group of order 2^m and of type $(1, 1, 1, \dots)$ and of a group which involves five and only five operators which are squares thereunder. Since such direct products have been excluded it results that the given subgroup of order 64 cannot be abelian.

Its commutator subgroup cannot include an operator of order 4 since

such an operator would be either t_1 or t_1^3 and hence it could not arise from an operator of order 4 or from an operator of order 2 contained in this subgroup. The operators of these two orders contained in this subgroup therefore generate a characteristic subgroup of order 32 under G . As each of the co-sets of this subgroup with respect to the subgroup formed by the squares under G involves 4 operators of order 2 and such an operator of order 2 cannot be transformed under this subgroup into itself multiplied by t_2^2 it results that the commutator subgroup of this group of order 64 is generated by t_1^2 . Its central is of order 16 and either of type $(2, 1, 1)$ or of type $(3, 1)$. In the former case it is easy to verify that no G can exist while in the latter case there is one such G . Hence *there are nine groups which have the property that each of them contains five and only five operators which are squares thereunder including at least one of order 4. Three of these are of order 32, five are of order 64, and one is of order 128.*

Suppose that the five operators which are squares under G include an operator of order 3 and hence two such operators. Since all the operators which are squares under G are relatively commutative² it results that such a G involves two and only two operators of order 2 which are squares and that each of the operators of order 4 in G transforms its operators of order 3 into their inverses. Hence such a G contains a subgroup of index 2 which is the direct product of its subgroup of order 3 and an abelian group of order 2^m and of type $(1, 1, 1, \dots)$. Each of its remaining operators is of order 4 since its operators of order 4 have two and only two distinct squares and every two operators of order 4 in G which have distinct squares are non-commutative. It therefore results that the commutator subgroup of G is the cyclic group of order 6 and that *there is one and only one group which satisfies the condition that it has five and only five operators which are squares including an operator of order 3. The order of this group is 48 and it contains the direct product of the group of order 3 and the abelian group of order 8 and of type $(1, 1, 1)$.*

It remains to consider the possible cases when the five operators which are squares include exactly four (s_1, s_2, s_3, s_4) which are of order 2. These four operators generate an abelian group whose order is either 8 or 16. We shall first prove that this order cannot be 16. If s_1, s_2, s_3, s_4 would generate a group H of order 16 then H would appear in the central of G for reasons which follow. Such an operator s_1 could not be non-commutative with another operator s_5 of order 2 contained in G for if it were s_1 and s_5 would generate

² *Ibid.*, vol. 19 (1933), pp. 1054-1057.

the octic group which would involve two of the three operators s_2, s_3, s_4 since the conjugate of an operator which is a square has the same property. As s_1 and these two operators would generate the four group the four operators s_1, s_2, s_3, s_4 could not then generate a group of order 16.

If one of these four operators s_2 would be non-commutative with an operator t_1 of order 4 contained in G , it may be assumed that $t_1^2 = s_1$ and that t_1 transforms s_2 and s_3 into each other and is therefore commutative with s_2s_3 . The group of order 8 generated by s_1, s_2, s_3 would be invariant under t_1 and $(t_1s_2)^2$ would equal $s_1s_2s_3$, which is impossible if s_1, s_2, s_3, s_4 generate a group of order 16. It therefore follows that if H were of order 16 it would appear in the central of G and every two operators of order 4 contained in G which have different squares would be non-commutative. If such a G exists we may assume without loss of generality that $t_1^2 = s_1, t_2^2 = s_2, t_3^2 = s_3$, and $t_4^2 = s_4$. It results directly that t_2 transforms t_1 into itself multiplied by one of the following five operators $s_1, s_2, s_1s_2, s_1s_2s_3, s_1s_2s_4$ since H includes the commutator subgroup of G . We shall first prove that the fourth of these cases is impossible and hence the fifth is also impossible.

In the fourth case t_1 and t_2 together with H generate a group of order 64 and t_1t_2 may be assumed to be t_3 . The operator t_4 transforms each of the operators t_1, t_2, t_1t_2 into itself multiplied respectively by one of the following operators: $s_1, s_4, s_1s_4, s_1s_4s_2, s_1s_4s_3; s_2, s_4, s_2s_4, s_2s_4s_1, s_2s_4s_3; s_3, s_4, s_3s_4, s_3s_4s_1, s_3s_4s_2$. This is impossible because t_4 transforms t_1t_2 into itself multiplied by the product of its two commutators with t_1 and t_2 . It therefore results that t_2 transforms t_1 into itself multiplied by one of the following three operators; s_1, s_2, s_1s_2 . If we can prove that the first of these is impossible it will also prove that the second is impossible. Hence we assume that the first condition is satisfied until we arrive at a contradiction.

The group of order 64 generated by H, t_1, t_2 then involves only operators of order 4 in addition to H . The operator t_3 gives rise to the following commutators with respect to t_1, t_2, t_1t_2 respectively: $s_1, s_3, s_1s_3; s_2, s_3, s_2s_3; s_2, s_3, s_2s_3$. Hence there is only one such subgroup of order 128 possible. In this t_3 gives rise to the following commutators s_3, s_2, s_2s_3 with respect to t_1, t_2, t_1t_2 respectively. Hence t_4 gives rise with respect to $t_1, t_2, t_1t_2, t_3, t_1t_3, t_2t_3, t_1t_2t_3$ respectively to the following commutators: $s_1, s_4, s_1s_4; s_2, s_4, s_2s_4; s_2, s_4, s_2s_4; s_3, s_4, s_3s_4; s_1, s_4, s_1s_4; s_3, s_4, s_3s_4; 1, s_1s_4, s_2s_4, s_3s_4$. As these are obviously inconsistent such a subgroup of order 128 cannot appear in G . The existence of such a G therefore implies that t_2 gives rise to the commutators s_1s_2 with respect to t_1 and hence it gives rise to the commutators s_2s_3, s_3s_4 with respect to t_3 and t_4 respectively. As this is impossible, since

it would give rise to too many squares, it has been proved that *when a group involves five and only five operators which are squares and four of them are of order 2 then these four operators generate an invariant subgroup of order 8.*

If an operator of order 2 in G would not be commutative with every operator of this invariant subgroup H then it would be non-commutative with one of its operators s_1 which is a square and it and s_1 would generate an octic group involving three operators which are squares. This operator would therefore be commutative with exactly half of the operators of H since it is commutative with the remaining square of order 2. Hence at least one of the squares of G would be invariant under G since the co-sets with respect to H are invariant and therefore two of these squares would have this property. An operator whose square is a non-invariant operator of G would therefore be commutative with every operator of H and hence all the operators of the co-set with respect to H to which it belongs have the same square. This square is therefore invariant under G , which is contrary to the hypothesis. That is, we arrived at a contradiction by assuming that an operator of order 2 in G is not commutative with every operator of H . If an operator of order 4 in G were non-commutative with a square it would transform exactly two squares among themselves. Since no operator could be non-commutative with all of the four squares these squares could not be transformed under G according to a group of degree 4. It has been noted that only two squares could not be non-invariant. Hence, it results that H is in the central of G .

It is easy to see that H is the central of G since this central cannot contain an operator of order 4 and if it would contain an operator of order 2 which does not appear in H then G would be a direct product. Each of the possible groups appears in one and only one of the following three categories: The first is composed of those in which the product of every two distinct squares of order 2 is a non-square, the second of those in which at least two operators of order 4 which have different squares are commutative, the third of those in which the product of two squares is a square but no two operators of order 4 which have different squares are commutative. In the first case it may be assumed that the four squares of order 2 are $s_1, s_2, s_3, s_1s_2s_3$. In each of the other two cases it may be assumed that the squares of order 2 are s_1, s_2, s_1s_2, s_3 . The smallest order of a group which satisfies the conditions under consideration is 64. We proceed to determine all the groups of this order which belong to these three categories in the given order.

We shall first prove that each of these groups contains a definite subgroup of order 32. To prove this we first extend H by an operator t_1 of order 4 whose square is s_1 and thus obtain an abelian subgroup of type $(2, 1, 1)$.

This subgroup is then extended by t_2 whose square is s_2 so as to obtain a subgroup of order 32. We proceed to prove that this can always be so selected that $t_1 t_2$ is of order 2 and hence it is completely determined. If $t_1 t_2$ were of order 4 its square may be assumed to be different from s_3 . Hence t_3 whose square is s_3 would have to transform this subgroup of order 32 so as to give rise to four commutators including at least one of the form $s_1 s_2$. An operator of order 4 with respect to which t_3 gives rise to this commutator and t_3 have a product of order 2 since the square of such a product could not be of the form $s_1 s_2$. This proves the following theorem: *If a group involves five and only five operators which are squares including four of order 2 and if the product of no two distinct ones of these operators of order 2 is a square then the group contains a subgroup of order 32 generated by these squares and two operators of order 4 having distinct squares and a product of order 2.*

It may now be assumed that all the groups of order 64 belonging to the first of the three categories under consideration contain this subgroup of order 32 generated by H, t_1, t_2 where $t_1 t_2$ is of order 2. There is obviously one and only one such group in which the product of every two operators of order 4 which have different squares is of order 2. There is also one and only one such group in which $t_1 t_3$ is of order 2 but $t_2 t_3$ is not of this order. To prove that in each of the remaining groups of this category each of the additional operators is of order 4 it is only necessary to note that $t_1 t_2 t_3$ could not be of order 2 in some one of them. This results from the fact that we may assume that $t_2 t_3$ is not of order 2 since we would otherwise get a group which is conjugate with the one already considered, and hence no commutator of the form $s_1 s_2$ could arise from t_3 . Each of these remaining groups therefore contains three pairs of operators of order 4 whose products are of order 2 and which are distinct modulo H and have distinct squares. In particular, each of these remaining groups involves three conjugate subgroups of order 32 which have the abelian subgroup of type $(1, 1, 1, 1)$ in common.

Since an operator whose square is $s_1 s_2 s_3$ appears among those which are added to the given group of order 32 it may be assumed without loss of generality that t_3 transforms t_1 into itself multiplied by s_2 and that it transforms $t_1 t_2$ into itself multiplied by one of the following four operators: 1, $s_1 s_2$, $s_1 s_3$, $s_2 s_3$. Hence there are two additional such groups of order 64. In one of these the commutator subgroup is of order 4 while in the other it is of order 8. It therefore results that *there are four and only four groups of order 64 which separately satisfy the condition that they contain five and only five operators which are squares, including four of order 2, and that the product of no two squares is a square.*

The second category of groups under consideration contains by hypothesis the abelian group of type $(2, 2, 1)$. To extend this so as to obtain a G of order 64 it is necessary to add thereto an operator t_4 of order 4 whose square s_4 is the fourth square of order 2 in G . Since t_4 is not commutative with any operator of order 4 contained in this subgroup it must give rise to four distinct commutators with respect thereto and hence it transforms into its inverse at least one of these operators of order 4. Since H contains three subgroups of order 4 which include the square of the operator of order 4 which is transformed into its inverse by t_4 and one of these subgroups corresponds to two possible G 's there results the following theorem: *Four and only four groups of order 64 have the property that each of them contains five operators which are squares thereunder and contains the abelian group of type $(2, 2)$.*

It remains to determine the groups of order 64 which separately satisfy the condition that no two of their operators of order 4 which have different squares are commutative but that the product of the squares of two such operators is one of the four squares of order 2 contained therein. We may assume that t_1 and t_2 are non-commutative and that $t_4^2 = s_1s_2$. We shall first consider the case when t_1t_2 is of order 2 and extend the subgroup of order 32 generated by H, t_1, t_2 by t_4 so as to obtain one of the groups of order 64 which satisfies the given condition. It is easy to verify that t_4 cannot transform one of the three operators t_1, t_2, t_1t_2 into itself multiplied by s_3 . It can also not transform t_1t_2 into itself multiplied by one of the following operators: $s_1s_2, s_1s_3, 1, s_1, s_2$ since its product with one of the three operators t_1, t_2, t_1t_2 has s_3 for its square. If t_1t_2 is transformed into itself multiplied by s_1s_2 the group is completely determined. The commutator subgroup of this group is of order 4.

When t_1t_2 is transformed into itself multiplied by $s_1s_2s_3$ the commutator subgroup is of order 8. There is one and only one such group and hence there are two possible groups of order 64 in which t_1t_2 is of order 2. When t_1t_2 is of order 4 the square of their product may be one of the two operators s_1, s_2 or it may be s_3 . In the former case we may assume without loss of generality that the square of t_1t_2 is s_2 . We do not need to consider the case when t_4 transforms t_1t_2 into itself multiplied by s_1 since G would then contain two operators of order 4 having different squares whose product would be of order 2. As before t_4 could not transform one of the operators t_1, t_2, t_1t_2 into itself multiplied by s_3 . Moreover, t_4 could not transform t_1t_2 into itself multiplied by one of the following operators: $1, s_2, s_1s_2s_3, s_2s_3$. If it transforms it into itself multiplied by s_1s_2 then G is completely determined and involves only

operators of order 4 in addition to H . Hence *there are seven and only seven groups of order 64 which separately satisfy the conditions that each contains five and only five operators which are squares, four being of order 2, and that no two operators of order 4 which have different squares are commutative but the product of two squares is a square.*

There is no upper limit for the orders of the remaining possible groups since every such group can be extended so as to obtain a group whose order is four times the order of the given group provided this group contains a subgroup of index 2 such that each of the remaining operators is of order 4. It is easy to verify that this condition is satisfied by groups in each of the three given categories and that when it is satisfied we may use the direct product of this group and a group of order 2 and adjoin to it an operator of order 4 which is commutative with an operator of order 4 not contained in the given subgroup of index 2, has the same square as the latter operator, and transforms this subgroup in the same manner as the given operator of order 4 transforms it. The resulting group can then be used in the same way to construct such a group of four times its own order and this process can be repeated indefinitely. Each of the infinite systems of groups thus obtained contains exactly five operators which are squares thereunder and four of these operators are of order 2.

CORRECTION AND ADDITION TO "COMPLEMENTS OF POTENTIAL THEORY."¹

By GRIFFITH C. EVANS.

Dr. F. G. Dressel has called my attention by means of an example to the necessity of a correction for Lemma II, p. 217, in the above mentioned memoir. In fact, the theorem of Daniell, quoted in the lemma, does not apply. The lemma should read as follows:

LEMMA II. *Let $f(x)$ be bounded and measurable in the Borel sense, $g_1(x), g_2(x)$ of bounded variation and $g_1(x)$ or $g_2(x)$ continuous, $a \leq x \leq b$; then $g(x) = g_1(x)g_2(x)$ is of bounded variation, and*

$$(1.1) \quad \int_a^b f(x) dg(x) = \int_a^b f(x) g_1(x) dg_2(x) + \int_a^b f(x) g_2(x) dg_1(x).$$

In the proof of the lemma omit (c), line 7, p. 218, replacing it by (d), and replace line 22, p. 218, by the inequality

$$t_m(b) - t_m(a) \leq t(b) - t(a),$$

for the total variation functions. Let $g_2(x)$, say, be continuous. There is then no need of $g_{2m}(x)$, and the proof, in the case of $f(x)$ continuous, ends with line 3, p. 219. The extension to $f(x)$, bounded and measurable in the Borel sense, is the same as before.

We note also the following:

LEMMA II'. *Let $g_1(x), g_2(x)$ be of bounded variation, $a \leq x \leq b$, t_1 the total variation of g_1 , and N_2 the upper bound of $|g_2(x)|$ over $a \leq x \leq b$. Then*

$$\left| \int_a^b g_1(x) dg_2(x) \right| \leq |g(b) - g(a)| + t_1 N_2.$$

¹ *American Journal of Mathematics*, vol. 54 (1932), pp. 213-234. The example given by Dr. Dressel is the following:

$$\begin{array}{ll} f(x) = 1, & 0 \leq x \leq 2, \\ g_1(x) = g_2(x) = 0, & 0 \leq x \leq 1, \\ & = 1, \quad 1 < x \leq 2, \end{array}$$

for which the left-hand member of (1.1) has the value 1 and the right-hand member the value 0.

In fact, let $g_{1n}(x)$ be the continuous polygonal approximation to $g_1(x)$, of Lemma II. Then, $g_{1n}(b)g_2(b) - g_{1n}(a)g_2(a) = g(b) - g(a)$, and

$$\begin{aligned}\int_a^b g_{1n}(x) dg_2(x) &= g(b) - g(a) - \int_a^b g_2(x) dg_{1n}(x) \\ \left| \int_a^b g_{1n}(x) dg_2(x) \right| &\leq |g(b) - g(a)| + \{t_{1n}(b) - t_{1n}(a)\}N_2 \\ &\leq |g(b) - g(a)| + t_1 N_2.\end{aligned}$$

But

$$\lim_{n=\infty} \int_a^b g_{1n}(x) dg_2(x) = \int_a^b g_1(x) dg_2(x),$$

so that the inequality is established.

In the proof of Lemma IV, p. 222, Lemma II' should be cited for the inequality of line 11, p. 224, without making use of the previous equation.

A general theorem of the type of Lemma II is the following one.

THEOREM. *Let $f(x)$ be bounded and measurable Borel, and $g_1(x)$, $g_2(x)$ of bounded variation, $a \leq x \leq b$; and let e_1 , e_2 be the sets respectively of values of x corresponding to the points of discontinuity of $g_1(x)$, $g_2(x)$. Then (1.1) is valid provided e_1 and e_2 have no points in common.*

There is evidently no loss in generality in assuming $g_1(x)$, $g_2(x)$ to be not negative and monotone-increasing. The function $g(x)$ will then be of the same sort. Let $f(x)$ be continuous, and write

$$g_i(x) = \alpha_i(x) + \beta_i(x), \quad (i = 1, 2),$$

where $\alpha_i(x)$ is continuous and $\beta_i(x)$ is the corresponding "function of discontinuities." We have

$$\int_a^b f(x) dg(x) = \int_a^b f d(\alpha_1 \alpha_2) + \int_a^b f d(\alpha_1 \beta_2) + \int_a^b f d(\alpha_2 \beta_1) + \int_a^b f d(\beta_1 \beta_2),$$

and by Lemma II, the identity (1.1) may be applied to every integral except the last. By proving that it applies to the last integral, the identity will be established for $f(x)$ continuous, and may then be extended to $f(x)$ bounded and measurable Borel, as before.

Accordingly it remains to prove that

$$\int_a^b f d(\beta_1 \beta_2) = \int_a^b f \beta_1 d\beta_2 + \int_a^b f \beta_2 d\beta_1,$$

assuming $f(x)$ to be continuous.

Consider first the case where β_1 and β_2 are step functions with merely a finite number of jumps B_i, C_j at values $x = b_i, x = c_j$ respectively, with $b_i \neq c_j$ for all i, j . Then, evidently,

$$\begin{aligned}\int_a^b f d(\beta_1 \beta_2) &= \sum_i f(b_i) \beta_2(b_i) B_i + \sum_j f(c_j) \beta_1(c_j) C_j \\ &= \int_a^b f(x) \beta_2(x) d\beta_1(x) + \int_a^b f(x) \beta_1(x) d\beta_2(x),\end{aligned}$$

where the integrals of the right-hand member are general (Daniell) integrals.

Let now $\beta_1(x)$ and $\beta_2(x)$ be arbitrary functions of discontinuities, so that they may have a denumerable infinity of finite jumps. We define $\beta_{1n}(x)$ as a step function, approximating to $\beta_1(x)$, with merely a finite number of jumps. In fact, let b_1, b_2, \dots, b_{k_n} be the values of x at which the jump of $\beta_1(x)$ is $\geq 1/n$. It may happen that a or b is a b_i . The function $\beta_{1n}(x)$ is to have discontinuities only at b_1, \dots, b_{k_n} and at these points it is to have the same discontinuities as $\beta_1(x)$, viz.,

$$\begin{aligned}\beta_{1n}(a) &= \beta_1(a) \\ \beta_{1n}(b_i) - \beta_{1n}(b_i \pm 0) &= \beta_1(b_i) - \beta_1(b_i \pm 0).\end{aligned}$$

It is clear that

$$(I) \quad \lim_{n \rightarrow \infty} \beta_{1n}(x) = \beta_1(x), \quad a \leq x \leq b.$$

For on the one hand, $\beta_{1n}(x) \leq \beta_1(x)$. And on the other hand, given $\epsilon > 0$, we can find n so that the sum of all the finite jumps which are each in value $< 1/n$, will be $< \epsilon$. Consequently for n sufficiently large, $\beta_{1n}(x) > \beta_1(x) - \epsilon$, $a \leq x \leq b$. Moreover, since $\beta_{1n}(x_2) - \beta_{1n}(x_1)$ is merely the sum of discontinuities of $\beta_1(x)$ belonging to those finite jumps of $\beta_1(x)$, in the interval $x_1 \leq x \leq x_2$, each of which is in value $\geq 1/n$, we have

$$(II) \quad \beta_{1n}(x_2) - \beta_{1n}(x_1) \leq \beta_1(x_2) - \beta_1(x_1), \quad a \leq x_1 < x_2 \leq b.$$

We define similarly functions $\beta_{2m}(x)$ approximating to $\beta_2(x)$, and similar properties (I), (II) hold for the functions $\beta_{mn}(x) = \beta_{1n}(x) \cdot \beta_{2m}(x)$. In fact,

$$\begin{aligned}\beta_{1n}(x_2) \beta_{2m}(x_2) - \beta_{1n}(x_1) \beta_{2m}(x_1) &= \{\beta_{1n}(x_2) - \beta_{1n}(x_1)\} \beta_{2m}(x_2) + \beta_{1n}(x_1) \{\beta_{2m}(x_2) - \beta_{2m}(x_1)\} \\ &\leq \{\beta_1(x_2) - \beta_1(x_1)\} \beta_2(x_2) + \beta_1(x_1) \{\beta_2(x_2) - \beta_2(x_1)\} \\ &= \beta_1(x_2) \beta_2(x_2) - \beta_1(x_1) \beta_2(x_1).\end{aligned}$$

The hypotheses (I), (II) are the hypotheses of Daniell's theorem,² whence

$$\lim_{m \rightarrow \infty} \int_a^b f(x) \beta_{1n}(x) d\beta_{2m}(x) = \int_a^b f(x) \beta_{1n}(x) d\beta_2(x).$$

But also

$$\lim_{n \rightarrow \infty} \int_a^b f(x) \beta_{1n}(x) d\beta_2(x) = \int_a^b f(x) \beta_1(x) d\beta_2(x),$$

as an elementary property of the general integral. Since e_1, e_2 have no common elements, there are no common points of discontinuity of β_{1n}, β_{2m} , and

$$\int_a^b f d\beta_{mn} = \int_a^b f \beta_{1n} d\beta_{2m} + \int_a^b f \beta_{2m} d\beta_{1n}.$$

By successive passages to the limit, the identity is established.

UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIFORNIA.

² P. J. Daniell, "Further properties of the general integral," *Annals of Mathematics*, vol. 21 (1920), pp. 203-220. See p. 218.

ON A CERTAIN CLASS OF ORTHOGONAL POLYNOMIALS.

By A. TARTLER.

Introduction. Let $\psi(x)$ denote a bounded non-decreasing function—"characteristic function"—with infinitely many points of increase on the finite or infinite interval (a, b) and such that

all "moments" $\alpha_i = \int_a^b x^i d\psi(x)$ exist ($i = 0, 1, \dots$), with $\alpha_0 > 0$.¹

The object of this paper is to study the system of orthogonal and normal polynomials

$$(1) \quad u_n(x; d\psi) \equiv u_n(x) \equiv \bar{a}_n(x^n - S_n x^{n-1} + \dots) \equiv \bar{a}_n U_n(x; d\psi) \equiv \bar{a}_n U_n(x) \\ (n = 0, 1, \dots; \bar{a}_n > 0)$$

corresponding to the more general characteristic function of bounded variation

$$\dot{\psi}(x) = \int_a^x (t - \alpha) d\psi(t) \quad (a < \alpha < b), \text{ with moments } \beta_i = \int_a^b x^i d\dot{\psi}(x). \\ (i = 1, 2, \dots),$$

and their relation to the system of polynomials

$$(2) \quad \phi_n(x) \equiv \phi_n(x; d\psi) \equiv a_n x^n + \bar{a}_{n,n-1} x^{n-1} + \dots \quad (a_n > 0; n = 0, 1, \dots) \\ \equiv a_n \Phi_n(x) \equiv a_n (x^n - S_n x^{n-1} + \dots),$$

having the fundamental property

$$(3) \quad \int_a^b \phi_m(x) \phi_n(x) d\psi(x) = \delta_{m,n} \quad (m, n = 0, 1, \dots).$$

(3) is equivalent to

$$(4) \quad \int_a^b \phi_n(x) G_{n-1}(x) d\psi(x) = 0,$$

where $G_s(x) = \sum_{i=0}^s g_i x^i$ here and hereafter stands for an arbitrary polynomial

¹ We assume the non-existence of numbers c, d such that

$$\int_a^c d\psi(x) = \int_d^b d\psi(x) = 0, \quad (a < c, d < b).$$

of degree $\leq s$. Our main purpose is to show how far the known properties of the system (2) are extendable to (1). The results obtained are an extension of those announced, without proof, by J. Shohat.²

1. *Some needed properties of orthogonal polynomials.*

$$(5) \quad \begin{cases} \text{(i)} & \left\{ \begin{aligned} \Phi_n(x) &= (x - c_n)\Phi_{n-1}(x) - \lambda_n\Phi_{n-2}(x); \quad \lambda_n = a_{n-2}^2/a_{n-1}^2; \\ c_n &= S_n - S_{n-1} = \int_a^b x\phi_{n-1}^2(x)d\psi(x) & (n \geq 2); \\ c_1 &= S_1 = \alpha_1/\alpha_0, \quad \lambda_1 = \alpha_0; \quad s_n = \sum_{i=1}^n c_i & (n \geq 1). \end{aligned} \right. \end{cases}$$

(ii) The polynomials $\Phi_n(x)$ are the denominators of the successive convergents of the "associated" continued fraction³

$$(6) \quad \int_a^b \frac{d\psi(y)}{x-y} \sim \frac{\lambda_1}{x-c_1} - \frac{\lambda_2}{x-c_2} - \dots$$

(iii) If $x_{1,n}, \dots, x_{n,n}$ denote the zeros of $\Phi_n(x)$, then

$$(7) \quad a < x_{1,n+1} < x_{1,n} < x_{2,n+1} < x_{2,n} < \dots < x_{n,n} < x_{n+1,n+1} < b.$$

(iv) Darboux' formula,⁴ which are of fundamental importance in the discussion which follows

$$(8) \quad \begin{cases} K_n(x, t; d\psi) \equiv K_n(x, t) \equiv \sum_{i=0}^n \phi_i(x)\phi_i(t) \\ \quad = \frac{a_n}{a_{n+1}} \frac{\phi_{n+1}(x)\phi_n(t) - \phi_{n+1}(t)\phi_n(x)}{x-t} \\ K_n(x, x; d\psi) \equiv K_n(x) \equiv \sum_{i=0}^n \phi_i^2(x) \\ \quad = \frac{a_n}{a_{n+1}} [\phi'_{n+1}(x)\phi_n(x) - \phi_{n+1}(x)\phi'_n(x)]. \end{cases}$$

(v) Let $\{\eta_{i,n}\}$, $\{\xi_{i,n}\}$ ($i=1, 2, \dots, n$) denote respectively the zeros of $\phi_n(x; d\psi_1)$, $\phi_n(x; d\psi_2)$, with

$$(9) \quad \begin{aligned} \psi_1(x) &\equiv \int_a^x (t-a)d\psi(t), \quad \psi_2(x) \equiv \int_a^x (b-t)d\psi(t) \quad (a \leq x \leq b): \\ a &< x_{1,n+1} < \xi_{1,n} < x_{1,n} < \eta_{1,n} < x_{2,n+1} < \xi_{2,n} < x_{2,n} < \eta_{2,n} \\ &< x_{3,n+1} < \dots < x_{n,n+1} < \xi_{n,n} < x_{n,n} < \eta_{n,n} < x_{n+1,n+1} < b. \end{aligned}$$

² Jacques Chokhate (J. Shohat), "Sur les fractions continues algébriques," *Comptes Rendus*, vol. 191 (1930), p. 474.

³ O. Perron, *Die Lehre von den Kettenbrüchen*, Teubner, 1913, p. 377.

⁴ Darboux, "Mémoire sur l'approximation des fonctions de très grands nombres," *Journal de Mathématiques* (3), vol. 4 (1878), pp. 5-56, 377-416.

2. *Existence of the system of orthogonal polynomials $U_n(x; d\psi)$.* The fundamental problem is to derive conditions assuring the existence of a sequence of polynomials $U_n(x)$ defined in (1), satisfying either one of the equivalent conditions of orthogonality:

$$(10) \quad \begin{cases} \int_a^b U_n(x) G_{n-1}(x) d\psi(x) = \int_a^b U_n(x) G_{n-1}(x) (x - \alpha) d\psi(x) = 0 \\ \int_a^b U_n(x) U_m(x) (x - \alpha) d\psi(x) = 0 \quad (n \neq m; n, m = 0, 1, 2, \dots). \end{cases} \quad (n = 1, 2, \dots),$$

Using (3, 10), we get, writing $(x - \alpha)U_n(x) = \sum_{i=0}^{n+1} A_i \phi_i(x)$:

$$(11) \quad \begin{aligned} (x - \alpha)U_n(x) &= A_{n+1}\phi_{n+1}(x) + A_n\phi_n(x), \\ A_{n+1} &= \frac{1}{a_{n+1}}, \quad A_n\phi_n(\alpha) = -\frac{\phi_{n+1}(\alpha)}{a_{n+1}}. \end{aligned}$$

If $\phi_n(\alpha) \neq 0$, A_n is uniquely determined. If $\phi_n(\alpha) = 0$, then necessarily $\phi_{n+1}(\alpha) \neq 0$ (see (7)) and A_n does not exist. This, combined with Darboux' formula (8), leads to

THEOREM I. *A necessary and sufficient condition that $U_n(x)$ satisfying (10) exist for a given n , is: $\phi_n(\alpha) \neq 0$. $U_n(x)$ is then uniquely determined:*

$$(12) \quad U_n(x) \equiv \frac{1}{a_{n+1}\phi_n(\alpha)} \frac{\phi_{n+1}(x)\phi_n(\alpha) - \phi_{n+1}(\alpha)\phi_n(x)}{x - \alpha} \equiv \frac{K_n(x, \alpha)}{a_n\phi_n(\alpha)}.$$

COROLLARY. *If $U_n(x)$ exist, then $U_n(\alpha) = \frac{K_n(\alpha)}{a_n\phi_n(\alpha)} \neq 0$.*

If $\phi_{n+1}(\alpha) = 0$, then by (11) $A_n = 0$ and

$$(13), (13') \quad U_n(x) \equiv \frac{\phi_{n+1}(x)}{a_{n+1}(x - \alpha)} \cdot \int_a^b U_n(x) G_n(x) d\psi(x) = 0;$$

i. e., the degree of the arbitrary polynomial $G_n(x)$ being here as high as n — the degree of $U_n(x)$. Conversely, by (12, 3), if a polynomial $U_n(x) \equiv x^n + \dots$ satisfies (13'), then necessarily $\phi_{n+1}(\alpha) = 0$. As an immediate consequence of Theorem I we state the important

THEOREM II. *$\phi_n(\alpha) \neq 0$ for $n \geq 1$ ⁵ implies the existence of a set of orthogonal polynomials $U_n(x) = x^n + \dots$ of all degrees ($n = 0, 1, \dots$) satisfying (10) and uniquely determined by means of (12).*

Hereafter we assume $\phi_n(\alpha) \neq 0$ ($n \geq 1$) unless explicitly stated otherwise.

⁵ Infinitely many such α exist in any subinterval of (a, b) .

3. *Normalization of the system $U_n(x)$.* By virtue of (8, 8)

$$\int_a^b K_n(x, \alpha) G_n(x) d\psi(x) = -\frac{\phi_{n+1}(\alpha) g_n}{a_{n+1}}.$$

Take here $G_n(x) \equiv K_n(x, \alpha)$ and use (12):

$$(14) \quad \begin{aligned} \int_a^b K_n^2(x, \alpha) d\psi(x) &= -\frac{a_n}{a_{n+1}} \phi_{n+1}(\alpha) \phi_n(\alpha), \\ \int_a^b U_n^2(x) d\psi(x) &= -\frac{\phi_{n+1}(\alpha)}{a_n a_{n+1} \phi_n(\alpha)} \equiv \frac{\rho_n}{\bar{a}_n^2} \quad (\rho_n = \pm 1; \bar{a}_n > 0), \\ \frac{1}{\bar{a}_n^2} &\equiv \left| \frac{\phi_{n+1}(\alpha)}{\phi_n(\alpha)} \right| \frac{1}{a_n a_{n+1}} \end{aligned}$$

$$(15) \quad \int_a^b u_n^2(x) d\psi(x) = \rho_n = \operatorname{sgn} \left[-\frac{\phi_{n+1}(\alpha)}{\phi_n(\alpha)} \right].$$

Thus the integral in (15) is positive or negative, contrary to what we shall call the "ordinary" case, i. e., that of a *monotonic* characteristic function. Turning to (7), we get at once

$$(16) \quad \begin{aligned} \rho_n &= +1 \quad \text{for } \alpha < x_{1,n+1}, \text{ or } x_{k,n} < \alpha < x_{k+1,n+1}; \\ \rho_n &= -1 \quad \text{for } x_{k,n+1} < \alpha < x_{k,n}, \text{ or } x_{n+1,n+1} < \alpha. \end{aligned}$$

4. *The recurrence relation for $U_n(x)$.* Write

$$(17) \quad U_n(x) = (x - \bar{c}_n) U_{n-1}(x) + P_{n-2}(x) \quad (\bar{c}_n = \text{const.}),$$

where $P_{n-2}(x)$ is a polynomial of degree $\leq n-2$. Making use of (10), we get at once:

$$(18) \quad 0 = \int_a^b P_{n-2}(x) G_{n-3}(x) (x - \alpha) d\psi(x).$$

The degree of $P_{n-2}(x)$ cannot be less than $n-3$. For otherwise, we could take in (18) $G_{n-3}(x) \equiv (x - \alpha) P_{n-2}(x)$ and thus render the integrand non-negative. $P_{n-2}(x)$ cannot be of degree $n-3$ for then (18) would be equivalent to (13'), which in turn implies $\phi_{n-2}(\alpha) = 0$, contrary to our assumption (§ 2). Hence, $P_{n-2}(x)$ is actually of degree $n-2$. Moreover, (18) being nothing but the condition of orthogonality (10), $P_{n-2}(x)$ differs from $U_{n-2}(x)$ by a constant factor only, so that (17) becomes

$$(19) \quad U_n(x) = (x - \bar{c}_n) U_{n-1}(x) - \bar{\lambda}_n U_{n-2}(x) \quad (n \geq 2; \bar{c}_n, \bar{\lambda}_n = \text{const.}).$$

We thus obtain for $\{U_n(x)\}$ a recurrence relation precisely of the same type as (5). (19) yields through (10, 14) by comparing coefficients:

$$(20) \quad \left\{ \begin{aligned} \bar{\lambda}_n &= \frac{\int_a^b U_{n-1}(x) x^{n-1} d\psi(x)}{\int_a^b U_{n-2}(x) x^{n-2} d\psi(x)} = \rho_{n-1} \rho_{n-2} \frac{\bar{a}_{n-2}^2}{\bar{a}_{n-1}^2} & (n \geq 2), \\ \bar{c}_n &= \frac{\int_a^b x U_{n-1}^2(x) d\psi(x)}{\int_a^b U_{n-1}^2(x) d\psi(x)} = \rho_{n-1} \int_a^b x U_{n-1}^2(x) d\psi(x) & (n \geq 2; \bar{c}_1 = \beta_1/\beta_0), \\ \bar{s}_n &= \sum_{i=1}^n \bar{c}_i, \quad \bar{c}_n = \bar{s}_n - \bar{s}_{n-1}. \end{aligned} \right.$$

It follows that in the case under consideration $\bar{\lambda}_n$ are not all positive, contrary to the ordinary case.

Introduce, as in the ordinary case (Perron, l. c.), the "associated" power series and continued fraction

$$(21) \quad \int_a^b \frac{d\psi(y)}{x-y} \sim \sum_{i=0}^{\infty} \frac{\beta_i}{x^{i+1}} \sim \frac{\bar{\lambda}_1/}{/q_1(x)} - \frac{\bar{\lambda}_2/}{/q_2(x)} - \cdots (q_i(x)\text{-polynomials}).$$

The n -th convergent of the latter we denote by $P_n(x)/Q_n(x)$, $Q_n(x)$ being of degree μ_n ($n \geq 0$). Then, its fundamental property:

$$(22) \quad \int_a^b \frac{d\psi(y)}{x-y} - \frac{P_n(x)}{Q_n(x)} = \left(\frac{1}{x^{\mu_n + \mu_{n+1}}} \right)^6$$

leads to the orthogonality property for $Q_n(x)$:

$$\int_a^b Q_n(x) G_{\mu_{n-1}}(x) d\psi(x) = 0.$$

Hence, we may identify $Q_n(x)$ with $U_{\mu_n}(x)$, or with $U_{\mu_{n-1}}(x)$ according as $\phi_{\mu_n}(\alpha)$ is, or is not, zero. If $\phi_n(\alpha) \neq 0$ for $n \geq 1$, the degrees of the denominators of the successive convergents in (21) differ by one and all the $q_n(x)$ are of the first degree, as in (6) above:

$$(23) \quad \int_a^b \frac{d\psi(y)}{x-y} \sim \frac{\bar{\lambda}_1/}{/x - \bar{c}_1} - \frac{\bar{\lambda}_2/}{/x - \bar{c}_2} - \cdots (\bar{\lambda}_2, \dots, \bar{c}_1, \dots \text{ from (20)}).$$

We take $\bar{\lambda}_1 = \beta_0$, for, by (22),

$$\int_a^b \frac{d\psi(y)}{x-y} - \frac{\bar{\lambda}_1}{x - \bar{c}_1} = \left(\frac{1}{x^3} \right).$$

5. The zeros of $U_n(x)$ compared with those of $\phi_{n+1}(x)$. Denote the zeros of $U_n(x)$ by $\bar{x}_{i,n}$ ($i = 1, 2, \dots, n$; $n \geq 1$). By (12):

⁶ $(1/x^s)$ generally stands for $c_1/x^s + c_2/x^{s+1} + \dots$ ($c_1 \neq 0$).

$$(24) \quad (x_{i,n+1} - \alpha)(x_{i+1,n+1} - \alpha)U_n(x_{i,n+1})U_n(x_{i+1,n+1}) \\ = \frac{\phi_{n+1}^2(\alpha)}{a_{n+1}^2\phi_n^2(\alpha)}\phi_n(x_{i,n+1})\phi_n(x_{i+1,n+1}) < 0.$$

Considering the sign of the product of the first two factors in the left-hand member of (24), we readily arrive at

THEOREM III. *The interval $(x_{i,n+1}, x_{i+1,n+1})$ contains either no zeros, or one zero, of $U_n(x)$, according as α is, or is not, an interior point of it.*

Remark. (13) shows that if α is one of the zeros of $\phi_{n+1}(x)$, its remaining n zeros are precisely those of $U_n(x)$. This case was excluded and is mentioned here merely as a limiting case when α tends to a zero of $\phi_{n+1}(x)$ (Cf. § 6).

COROLLARY. *If $\alpha < x_{1,n+1}$ or $> x_{n+1,n+1}$, the zeros of $U_n(x)$ separate those of $\phi_{n+1}(x)$.*

We proceed to investigate more closely the case when α is an interior point of one of the intervals $(x_{k,n+1}, x_{k+1,n+1})$ ($k = 1, 2, \dots, n$). Since $x_{k,n+1} < x_{k,n} < x_{k+1,n+1}$, it is convenient to consider two cases:

(i) $x_{k,n+1} < \alpha < x_{k,n}$. In (12) put $x = x_{n+1,n+1}$:

$$(x_{n+1,n+1} - \alpha)U_n(x_{n+1,n+1}) = -\frac{1}{a_{n+1}}\frac{\phi_{n+1}(\alpha)}{\phi_n(\alpha)}\phi_n(x_{n+1,n+1}).$$

Here $x_{n+1,n+1} - \alpha > 0$, $-\frac{\phi_{n+1}(\alpha)}{\phi_n(\alpha)} < 0$ (see (7)), $\phi_n(x_{n+1,n+1}) > 0$; hence $U_n(x_{n+1,n+1}) < 0$ and $U_n(x)$ has one and only one zero in the interval $(x_{n+1,n+1}, +\infty)$.

(ii) $x_{k,n} < \alpha < x_{k+1,n+1}$. We find in a similar manner that $U_n(x)$ has one and only one zero in the interval $(-\infty, x_{1n+1})$.

We proceed further to specify the values of α for which the zeros of $U_n(x)$, for a given n , include either a or b (assumed to be finite). To this end consider $\Phi_n(x) \equiv \Phi_n(x; d\psi_2)$ for which

$$\int_a^b \left\{ \frac{(b-x)\Phi_n(x)}{x-\xi_{k,n}} \right\} G_{n-1}(x)(x-\xi_{k,n})d\psi(x) = 0 \quad (1 \leq k \leq n).$$

Hence, by virtue of Theorem I (uniqueness):

$$U_n(x; (x-\xi_{k,n})d\psi) \equiv -\frac{(b-x)\Phi_n(x; d\psi_2)}{x-\xi_{k,n}}.$$

Similarly we treat the point $x = a$ by means of $\Phi_n(x; d\psi_1)$. In other words, if α is a zero of the polynomial $\Phi_n(x; d\psi_{1,2})$, one of the zeros of $U_n(x)$ coincides correspondingly with a or b . This conclusion fully harmonizes with the inequalities and the results of §§ 2, 3.

6. The zeros $\{\bar{x}_{i,n}\}$ of $U_n(x)$ as functions of α .

THEOREM IV. $\{\bar{x}_{i,n}\}$ increase with α .

Proof. Differentiate with respect to α the relation $K_n(\bar{x}_{i,n}, \alpha) = 0$ (see (12)):

$$(25) \quad \frac{d\bar{x}_{i,n}}{d\alpha} = - \frac{\frac{\partial K_n(\bar{x}_{i,n}, \alpha)}{\partial \alpha}}{\frac{\partial K_n(\bar{x}_{i,n}, \alpha)}{\partial \bar{x}_{i,n}}} \quad (i = 1, 2, \dots, n; n \geq 1).$$

Develop the right-hand member in (25), making use of (12):

$$\frac{d\bar{x}_{i,n}}{d\alpha} = \frac{g(\alpha, \bar{x}_{i,n})}{g(\bar{x}_{i,n}, \alpha)} \quad (g(x, y) \equiv \phi'_{n+1}(x)\phi_n(y) - \phi_{n+1}(y)\phi'_n(x)).$$

The desired result, namely

$$\frac{d\bar{x}_{i,n}}{d\alpha} > 0 \quad (i = 1, 2, \dots, n; n \geq 1)$$

will be established if we succeed in showing that

$$g(\alpha, \bar{x}_{i,n})g(\bar{x}_{i,n}, \alpha) > 0.$$

But this latter inequality follows from the readily verifiable identity

$$\begin{aligned} g(x, y)g(y, x) &\equiv g(x)g(y) \\ &+ [\phi'_n(x)\phi'_{n+1}(y) - \phi'_n(y)\phi'_{n+1}(x)][\phi_{n+1}(x)\phi_n(y) - \phi_{n+1}(y)\phi_n(x)], \\ (g(x) &\equiv g(x, x) = \frac{a_{n+1}}{a_n} \sum_{i=0}^n \phi_i^2(x)) \end{aligned}$$

which leads to

$$g(\alpha, \bar{x}_{i,n})g(\bar{x}_{i,n}, \alpha) = g(\alpha)g(\bar{x}_{i,n}) = \frac{a_{n+1}^2}{a_n^2} \sum_{i=0}^n \phi_i^2(\alpha) \sum_{i=0}^n \phi_i^2(\bar{x}_{i,n}) > 0.$$

Remark. The above general theorem holds for any real α , inside or outside (a, b) .

The results of § 5, together with Theorem IV, are sufficient to describe completely the behavior of $\{\bar{x}_{i,n}\}$ as α varies increasingly from $-\infty$ to $+\infty$. This description is summarized in the following table. It will be recalled

(§ 2) that when α is a zero of $\phi_n(x)$, $U_n(x)$ does not exist, but $U_{n-1}(x)$ necessarily exists; at this point it is convenient to regard it as the polynomial $U_n(x)$, with one of its zeros infinite.

α	$\bar{x}_{i,n}$
$\alpha < x_{1,n+1}$	$x_{i,n+1} < \bar{x}_{i,n} < x_{i+1,n+1} \quad (i=1, 2, \dots, n)$
$* \alpha = x_{1,n+1}$	$\bar{x}_{i,n} = x_{i+1,n+1} \quad (i=1, 2, \dots, n)$
$x_{1,n+1} < \alpha < \xi_{1,n}$	$x_{i,n+1} < \bar{x}_{i-1,n} < x_{i+1,n+1} \quad (i=2, 3, \dots, n+1; x_{n+2,n+1} \equiv b)$
$\alpha = \xi_{1,n}$	$\bar{x}_{i,n} = \xi_{i+1,n} \quad (i=1, 2, \dots, n-1), \quad \bar{x}_{n,n} = b$
$\xi_{1,n} < \alpha < x_{1,n}$	$x_{i,n+1} < \bar{x}_{i-1,n} < x_{i+1,n+1} \quad (i=2, 3, \dots, n), \quad \bar{x}_{n,n} > b$
$\alpha = x_{1,n}$	$\bar{x}_{n,n} = +\infty$, or $\bar{x}_{1,n} = -\infty$
$x_{1,n} < \alpha < \eta_{1,n}$	$\bar{x}_{1,n} < a, x_{i,n+1} < \bar{x}_{i,n} < x_{i+1,n+1} \quad (i=2, 3, \dots, n)$
$\alpha = \eta_{1,n}$	$\bar{x}_{1,n} = a, \bar{x}_{i,n} = \eta_{i,n} \quad (i=2, 3, \dots, n)$
$\eta_{1,n} < \alpha < x_{2,n+1}$	$a < \bar{x}_{1,n} < x_{1,n+1}, x_{i,n+1} < \bar{x}_{i,n} < x_{i+1,n+1} \quad (i=2, 3, \dots, n)$
$** \alpha = x_{2,n+1}$	$\bar{x}_{1,n} = a, \bar{x}_{i,n} = \eta_{i,n} \quad (i=2, 3, \dots, n)$
α varies from $x_{2,n+1}$ to $x_{n+1,n+1}$	$\bar{x}_{i,n}$ varies as from $*$ to $**$, with proper changes of indices.
$x_{n+1,n+1} < \alpha$	$x_{i,n+1} < \bar{x}_{i,n} < x_{i+1,n+1} \quad (i=1, 2, \dots, n).$

7. On the separation of the zeros of $U_n(x)$ and $U_{n+1}(x)$. Here we use $K_n(x, \alpha)$ instead of $U_n(x)$. Consider first the case when $a < \alpha < x_{1,n+1}$. Here (§ 5) the zeros $\bar{x}_{i,n}$ of $K_n(x, \alpha)$ separate the zeros of $\phi_{n+1}(x)$:

$$(26) \quad a < x_{1,n+1} < \bar{x}_{1,n} < x_{2,n+1} < \bar{x}_{2,n} < \dots < \bar{x}_{n,n} < x_{n+1,n+1} < b.$$

We note that if (26) holds for $n = n_0$, it does so for $n < n_0$, for the hypothesis $\alpha < x_{1,n+1}$ implies $\alpha < x_{1,m}$, $m < n+1$, by virtue of (7). Furthermore,

$$(27) \quad K_{n+1}(\bar{x}_{i,n}, \alpha) K_{n+1}(\bar{x}_{i+1,n}, \alpha) = \phi_{n+1}^2(\alpha) \phi_{n+1}(\bar{x}_{i,n}) \phi_{n+1}(\bar{x}_{i+1,n}),$$

$$(28) \quad K_{n+1}(x_{i,n+1}, \alpha) K_{n+1}(x_{i+1,n+1}, \alpha) = K_n(x_{i,n+1}, \alpha) K_n(x_{i+1,n+1}, \alpha).$$

The right-hand member of (27) being negative by virtue of (26), it follows that $K_{n+1}(x, \alpha)$ changes sign an odd number of times in each of the intervals $(\bar{x}_{i,n}, \bar{x}_{i+1,n})$ ($i=1, 2, \dots, n-1$). Moreover, the right-hand member of (28) being also negative, we conclude that it changes sign in each of these intervals only once. Hence,

$$\alpha < x_{1,n+1} \text{ implies } \bar{x}_{1,n+1} < \bar{x}_{1,n} < \bar{x}_{2,n+1} < \bar{x}_{2,n} < \dots < \bar{x}_{n,n} < \bar{x}_{n+1,n+1}.$$

The same inequalities hold if $x_{n+1,n+1} < \alpha$.

Assume now that α separates two of the zeros of $\phi_{n+1}(x)$; say, $x_{k,n+1} < \alpha < x_{k+1,n+1}$. Here (§ 5) $K_n(x, \alpha)$ has no zero in $(x_{k,n+1}, x_{k+1,n+1})$ and has one zero in each of the remaining intervals $(x_{i,n+1}, x_{i+1,n+1})$ ($i \neq k$). If in (27) $i \neq k-1$, its right-hand member is negative, and $K_{n+1}(x, \alpha)$ changes sign in the corresponding intervals $(\bar{x}_{i,n}, \bar{x}_{i+1,n})$ ($i \neq k$) at least once. We now assert that $K_{n+1}(x, \alpha)$ changes sign twice in the interval $(\bar{x}_{k-1,n}, \bar{x}_{k,n})$; more precisely, it changes sign in each of the subintervals $(\bar{x}_{k-1,n}, x_{k,n+1})$, $(x_{k+1,n+1}, \bar{x}_{k,n})$. In fact,

$$K_{n+1}(\bar{x}_{k-1,n}, \alpha) = \phi_{n+1}(\bar{x}_{k-1,n}) \phi_{n+1}(\alpha),$$

$$K_{n+1}(x_{k,n+1}, \alpha) = K_n(x_{k,n+1}, \alpha) = -\frac{a_n}{a_{n+1}} \frac{\phi_{n+1}(\alpha) \phi_n(x_{k,n+1})}{x_{k,n+1} - \alpha},$$

$$K_{n+1}(\bar{x}_{k-1,n}, \alpha) K_{n+1}(x_{k,n+1}, \alpha) = -\frac{a_n}{a_{n+1}} \frac{\phi_{n+1}^2(\alpha) \phi_{n+1}(\bar{x}_{k-1,n}) \phi_n(x_{k,n+1})}{x_{k,n+1} - \alpha}.$$

Furthermore, since $\text{sgn } \phi_{n+1}(\bar{x}_{k-1,n}) = (-1)^{n+k}$, $\text{sgn } \phi_n(x_{k,n+1}) = (-1)^{n+k-1}$,

$$K_{n+1}(\bar{x}_{k-1,n}, \alpha) K_{n+1}(x_{k,n+1}, \alpha) < 0,$$

and similarly,

$$K_{n+1}(\bar{x}_{k,n}, \alpha) K_{n+1}(x_{k+1,n+1}, \alpha) < 0.$$

The last two inequalities prove our assertion. Thus we state

THEOREM V. (i) $\alpha < x_{1,n+1}$ or $\alpha > x_{n+1,n+1}$ implies: the zeros of $U_n(x)$ separate those of $U_{n+1}(x)$; (ii) $x_{k,n+1} < \alpha < x_{k+1,n+1}$ implies: each of the intervals $(\bar{x}_{i,n}, \bar{x}_{i+1,n})$ ($i \neq k-1$) contains one zero and the interval $(\bar{x}_{k-1,n}, \bar{x}_{k,n})$ contains two zeros of $U_{n+1}(x)$.

(The one remaining zero of $U_{n+1}(x)$ is either $< \bar{x}_{1,n}$ or $> \bar{x}_{n,n}$).

It is known for the ordinary case that the zeros of $\Phi_n(x)$ for n very large are everywhere dense in (a, b) , provided

$$(29) \quad \int_{a_1}^{b_1} d\psi(x) \neq 0 \quad (a \leq a_1 < b_1 \leq b)$$

This, combined with Theorem V leads to the

COROLLARY. Under (29) the zeros of $U_n(x)$ for n very large are everywhere dense in (a, b) .

8. *The mechanical quadratures formula related to $U_n(x)$.* We consider the mechanical quadratures formula—a direct application of the Lagrange interpolation formula—

$$(30) \quad \int_a^b G_{n-1}(x) d\psi(x) = \sum_{i=1}^n h_{i,n} G_{n-1}(\xi_i) \\ \left[h_{i,n} \equiv \int_a^b \frac{\prod_{i=1}^n (x - \xi_i) d\psi(x)}{(x - \xi_i) \left\{ \prod_{i=1}^n (x - \xi_i) \right\}'_{x=\xi_i}} \right],$$

where $\{\xi_i\}$ denote n distinct points arbitrarily chosen. If we write

$$G_{2n-1}(x) = G_{n-1}(x) \prod_{i=1}^n (x - \xi_i) + G_{n-1}^{(1)}(x)$$

and make use of the orthogonality properties (10), we arrive at

THEOREM VI. *The mechanical quadratures formula (30) holds for $G_{2n-1}(x)$ if, and only if, the points ξ_i are zeros of $U_n(x)$.*

We thus get a formula of Gauss' type

$$(31) \quad \int_a^b G_{2n-1}(x) d\psi(x) = \sum_{i=1}^n \bar{H}_{i,n} G_{2n-1}(\bar{x}_{i,n}) \\ \left(\bar{H}_{i,n} \equiv \int_a^b \frac{U_n(x) d\psi(x)}{(x - \bar{x}_{i,n}) U'_n(\bar{x}_{i,n})} \right).$$

We get further, knowing that $\bar{x}_{i,n} \neq \alpha$, and taking in (31) successively

$$G_{2n-1}(x) \equiv \left\{ \frac{U_n(x)}{(x - \bar{x}_{i,n}) U'_n(\bar{x}_{i,n})} \right\}^2, \quad \left\{ \frac{U_n(x)}{(x - \bar{x}_{i,n}) U'_n(\bar{x}_{i,n})} \right\}^2 (x - \alpha): \\ \bar{H}_{i,n} = \int_a^b \left\{ \frac{U_n(x)}{(x - \bar{x}_{i,n}) U'_n(\bar{x}_{i,n})} \right\}^2 (x - \alpha) d\psi(x), \\ (32) \quad \bar{H}_{i,n} = \frac{1}{\bar{x}_{i,n} - \alpha} \int_a^b \left\{ \frac{U_n(x)}{(x - \bar{x}_{i,n}) U'_n(\bar{x}_{i,n})} \right\}^2 (x - \alpha)^2 d\psi(x) \\ \begin{matrix} < 0 & \text{for } \bar{x}_{i,n} < \alpha \\ > 0 & \text{for } \bar{x}_{i,n} > \alpha \end{matrix}.$$

Here again we find an essential difference between the case under consideration and the ordinary one (where all the coefficients in the mechanical quadratures formula of type (31) are positive).

We proceed to derive an interesting expression for $\bar{H}_{i,n}$ in terms of $K_n(x)$. The orthogonality property (10) rewritten as

$$\int_a^b \left\{ \frac{U_n(x)(x - \alpha)}{x - \bar{x}_{i,n}} \right\} G_{n-1}(x) (x - \bar{x}_{i,n}) d\psi(x) = 0$$

shows the existence of a polynomial of degree n

$$U_n(x; (x - \bar{x}_{i,n}) d\psi) \equiv \frac{U_n(x; d\psi)(x - \alpha)}{x - \bar{x}_{i,n}},$$

orthogonal in (a, b) with respect to the characteristic function

$$\psi_{i,n} \equiv \int_a^x (t - \bar{x}_{i,n}) d\psi(t).$$

We derive successively:

$$U_n(x; d\psi) = \frac{U_n(x; d\psi_{i,n})(x - \bar{x}_{i,n})}{x - \alpha}; \quad U'_n(\bar{x}_{i,n}; d\psi) = \frac{U_n(\bar{x}_{i,n}; d\psi_{i,n})}{\bar{x}_{i,n} - \alpha}.$$

Substitute in (31) and apply (12, 8, 3):

$$(33) \quad \bar{H}_{i,n} = \frac{\bar{x}_{i,n} - \alpha}{K_n(\bar{x}_{i,n}; d\psi)} \quad (i = 1, 2, \dots, n).$$

(33) is another proof of the inequalities in (32). It also may give indication as to the asymptotic behavior of $\bar{H}_{i,n}$ for $n \rightarrow \infty$.

9. *Extension of Darboux' formulae.* The recurrence relation (19) readily leads to

$$\begin{aligned} \sqrt{\rho_{n-1}\rho_n\bar{\lambda}_{n+1}} \frac{u_n(x)u_{n-1}(y) - u_n(y)u_{n-1}(x)}{x - y} &= u_{n-1}(x)u_{n-1}(y) \\ &+ \sqrt{\rho_{n-2}\rho_{n-1}\bar{\lambda}_n} \frac{u_{n-1}(x)u_{n-2}(y) - u_{n-1}(y)u_{n-2}(x)}{x - y}, \\ \bar{K}_n(x, y) &\equiv \sum_{i=0}^n u_i(x)u_i(y) = \frac{\bar{a}_n}{\bar{a}_{n+1}} \frac{u_{n+1}(x)u_n(y) - u_{n+1}(y)u_n(x)}{x - y}, \\ \bar{K}_n(x, x) &\equiv \bar{K}_n(x) = \frac{\bar{a}_n}{\bar{a}_{n+1}} [u'_{n+1}(x)u_n(x) - u_{n+1}(x)u'_n(x)]. \end{aligned}$$

Thus Darboux' formulae (8) hold in our case without any modification.

10. *A mechanical quadratures formula with a fixed interior point.* Consider the mechanical quadratures formula

$$(34) \quad \int_a^b G_n(x) d\psi(x) = \sum_{i=0}^n H_{i,n} G_n(\xi_i)$$

$$\left[H_{i,n} \equiv \int_a^b \frac{\prod_{i=0}^n (x - \xi_i) d\psi(x)}{(x - \xi_i) \left\{ \prod_{i=0}^n (x - \xi_i) \right\}'_{x=\xi_i}} \right],$$

where the points $\{\xi_i\}$ ($i = 0, 1, \dots, n$) are distinct and $\xi_0 = \alpha$ is arbitrarily fixed inside (a, b) . We may show by the method of § 8 that (34) holds for $G_{2n}(x)$, provided $\phi_n(\alpha) \neq 0$ ($n = 1, 2, \dots$), $\xi_{i,n} = \bar{x}_{i,n}$ ($i = 1, 2, \dots, n$), the zeros of $U_n(x)$, so that (see (33))

^{*} This could be illustrated by means of Hermite polynomials.

$$(35) \quad H_{i,n} = \frac{\bar{H}_{i,n}}{\bar{x}_{i,n} - \alpha} = \frac{1}{K_n(\bar{x}_{i,n})} \quad (i = 1, 2, \dots, n), \quad H_{0,n} = \frac{1}{K_n(\alpha)}.$$

Assume (a, b) to be finite. Since $\bar{x}_{i,n}$ ($i = 2, 3, \dots, n-1$) and at least one of the two zeros $\bar{x}_{1,n}, \bar{x}_{n,n}$ are always in (a, b) , (35) shows that the corresponding $\bar{H}_{i,n}$ tend to zero as $n \rightarrow \infty$, in all cases for which it is known that $H_{i,n} \rightarrow 0$. (See § 11 below).

We get further, taking in (34) successively

$$\begin{aligned} G_{2n}(x) &\equiv \left\{ \frac{(x-\alpha)U_n(x)}{(x-\bar{x}_{i,n})(\bar{x}_{i,n}-\alpha)U'_n(\bar{x}_{i,n})} \right\}^2, \quad \left\{ \frac{U_n(x)}{U_n(\alpha)} \right\}^2, \quad 1; \\ H_{i,n} &= \int_a^b \left\{ \frac{(x-\alpha)U_n(x)}{(x-\bar{x}_{i,n})(\bar{x}_{i,n}-\alpha)U'_n(\bar{x}_{i,n})} \right\}^2 d\psi(x) \\ &\quad (i = 1, 2, \dots, n), \\ H_{0,n} &= \int_a^b \left\{ \frac{U_n(x)}{U_n(\alpha)} \right\}^2 d\psi(x), \\ (36) \quad \sum_{i=0}^n H_{i,n} &= \int_a^b d\psi(x). \end{aligned}$$

We see that here all $H_{i,n}$ are positive.

11. *Tchebycheff inequalities related to $U_n(x)$.* Denote the set of points $\alpha, \{\bar{x}_{i,n}\}$ ($i = 1, 2, \dots, n$) by $y_{1,n+1} < y_{2,n+1} < \dots < y_{n+1,n+1}$, change the numbering of $H_{i,n}$ in (34) accordingly and rewrite it as

$$\int_a^b G_{2n}(x) d\psi(x) = \sum_{i=1}^{n+1} H_i G_{2n}(y_{i,n+1}).$$

Following Stieltjes⁸ (and Markoff), construct $G_{2n}(x)$ subject to the following conditions:

$$\begin{aligned} G_{2n}(y_{i,n+1}) &= 1 & (i = 1, 2, \dots, k), \\ G_{2n}(y_{i,n+1}) &= 0 & (i = k+1, k+2, \dots, n+1), \\ G'_{2n}(y_{i,n+1}) &= 0 & (1 \leq i \leq n+1, i \neq k). \end{aligned}$$

These $2n+1$ conditions determine $G_{2n}(x)$ uniquely. Moreover, $G'_{2n}(x)$ has n zeros at the points $y_{i,n+1}$ ($i = 1, 2, \dots, k-1, k+1, \dots, n+1$) and, by Rolle's theorem, $k-1$ zeros inside $(y_{i,n+1}, y_{i+1,n+1})$ ($i = 1, 2, \dots, k-1$), and $n-k$ zeros inside $(y_{i,n+1}, y_{i+1,n+1})$ ($i = k, k+1, \dots, n+1$), with $n + (k-1) + n - k = 2n - 1$. It follows readily that

$$G_{2n}(x) \geq 0 \text{ for all } x, \geq 1 \text{ for } x \leq y_{k,n+1},$$

⁸ Stieltjes, "Quelques recherches sur les quadratures dites mécaniques," *Œuvres*, vol. 1, pp. 377-396.

$$\int_a^b G_{2n}(x) d\psi(x) = H_1 + H_2 + \cdots + H_k \geq \int_a^{y_{k,n+1}} G_{2n}(x) d\psi(x) \\ \begin{cases} k = 1, 2, \cdots, n+1, & \text{if } \bar{x}_{n,n} \leq b \\ k = 1, 2, \cdots, n, & \text{if } \bar{x}_{n,n} > b \end{cases}$$

$$(37) \quad H_1 + H_2 + \cdots + H_k \geq \int_a^{y_{k,n+1}} d\psi(x) \\ \begin{cases} k = 1, 2, \cdots, n+1, & \text{if } \bar{x}_{n,n} \leq b \\ k = 1, 2, \cdots, n, & \text{if } \bar{x}_{n,n} > b. \end{cases}$$

Similarly, using the polynomial $T_{2n}(x)$ such that

$$\begin{aligned} T_{2n}(y_{i,n+1}) &= 0 & (i = 1, 2, \cdots, k); \\ T_{2n}(y_{i,n+1}) &= 1 & (i = k+1, k+2, \cdots, n+1); \\ T'_{2n}(y_{i,n+1}) &= 0 & (1 \leq i \leq n+1; i \neq k+1); \end{aligned}$$

$$(38) \quad H_{k+1} + H_{k+2} + \cdots + H_{n+1} \geq \int_{y_{k+1,n+1}}^b d\psi(x) \quad (k = 1, 2, \cdots, n),$$

and combining this with (36):

$$(39) \quad H_1 + H_2 + \cdots + H_k \leq \int_a^{y_{k+1,n+1}} d\psi(x) \quad (k = 1, 2, \cdots, n).$$

The inequalities (37, 39) constitute an extension to our case of the important Tchebycheff inequalities. It follows readily that

$$(40) \quad H_k \leq \int_{y_{k-1,n+1}}^{y_{k+1,n+1}} d\psi(x) \quad (k = 2, 3, \cdots, n),$$

$$H_1 \leq \int_a^{y_{2,n+1}} d\psi(x), \quad H_{n+1} \leq \int_{y_{n,n+1}}^b d\psi(x).$$

Hence, if $\psi(x)$ is continuous in the finite interval (a, b) which contains no subinterval (a_1, b_1) such that $\int_{a_1}^{b_1} d\psi(x) = 0$, then (see Corollary to Theorem V) $H_i \rightarrow 0$ as $n \rightarrow \infty$ ($i = 1, 2, \cdots, n+1$). By virtue of (35) we infer that $K_n(\alpha) \rightarrow \infty$ as $n \rightarrow \infty$ for any fixed α . This result combined with a theorem due to Hamburger⁹ gives a direct and elementary proof of the important fact that the moment problem for a finite interval is determined.

12. On the associated continued fraction.

THEOREM VI. If $r(x)$ is a continuous function having s changes of sign between a and b and $\psi(x)$ is of the nature indicated, then in the associated continued fraction

$$(41) \quad F(x) \equiv \int_a^b \frac{r(y) d\psi(y)}{x-y} \sim \frac{\lambda_1/}{/q_1(x)} - \frac{\lambda_2/}{/q_2(x)} - \cdots$$

⁹ H. Hamburger, "Über eine Erweiterung des Stieltjesschen Momentenproblem," *Mathematische Annalen*, vol. 81 (1920), pp. 235-319, Theorem XVII.

the degrees of the polynomials $q_i(x)$ ($i = 1, 2, \dots$) cannot exceed $s + 1$.

Proof. We have formally, denoting the i -th convergent of (41) by $\Omega_i(x)/\Phi_i(x)$, with $\Phi_i(x) \equiv x^{\mu_i} + p_{\mu_i-1}x^{\mu_i-1} + \dots$:

$$F(x) = \sum_{n=0}^{\infty} \frac{\alpha_n}{x^{n+1}} \quad (\alpha_n \equiv \int_a^b r(y) y^n d\psi(y)), \quad F(x)\Phi_i(x) - \Omega_i(x) = \left(\frac{1}{x^{\mu_{i+1}}}\right),$$

and expanding the left-hand member:

$$\alpha_j p_0 + \alpha_{j+1} p_1 + \dots + \alpha_{j+\mu_i} = 0 \quad (j = 0, 1, \dots, \mu_{i+1} - 2),$$

which is equivalent to

$$(42) \quad \int_a^b r(x) \Phi_i(x) G_{\mu_{i+1}-2}(x) d\psi(x) = 0 \quad (i = 1, 2, \dots).$$

Were a certain $q_i(x)$ in (41) of degree $> s + 1$, we would have

$$\mu_{i+1} - \mu_i > s + 1, \quad \mu_{i+1} - 2 \geq \mu_i + s,$$

and we could render (42) impossible by choosing $G_{\mu_{i+1}-2}(x)$ in (42) so that $r(x)G_{\mu_{i+1}-2}(x) \geq 0$ for $a \leq x \leq b$. The results of § 13 below show that *the upper bound for the degree of $q_i(x)$ as given in Theorem VI is the best possible.*

13. *The case $\phi_n(\alpha) = 0$.* Here (§ 2) $U_n(x)$ does not exist, $U_{n-1}(x)$ and $U_{n+1}(x)$, however, necessarily exist (since $\phi_{n-1}(\alpha)\phi_{n+1}(\alpha) \neq 0$). Moreover, by (13'),

$$\int_a^b U_{n-1}(x) G_{n-1}(x) (x - \alpha) d\psi(x) = 0$$

We get, writing

$$(43) \quad U_{n+1}(x) = (x^2 + \bar{b}_{n+1}x + \bar{c}_{n+1})U_{n-1}(x) + P_{n-2}(x) \\ (P_{n-2}(x) - \text{polynomial of degree} \leq n-2):$$

$$(44) \quad \int_a^b P_{n-2}(x) G_{n-3}(x) (x - \alpha) d\psi(x) = 0 \quad (\text{see (10)}).$$

(44) can be satisfied in the following cases only

- (i) $P_{n-2}(x) \equiv 0$; (ii) $P_{n-2}(x) \equiv -l_{n+1}U_{n-2}(x)$, if $\phi_{n-2}(\alpha) \neq 0$;
- (iii) $P_{n-2}(x) \equiv -l_{n+1}U_{n-3}(x)$, if $\phi_{n-2}(\alpha) = \phi_n(\alpha) = 0$.

(i) is impossible. In fact, it leads, through (13', 10), to

$$U_{n+1}(x) = (x^2 + \bar{b}_{n+1}x + \bar{c}_{n+1})U_{n-1}(x), \quad \int_a^b x^n U_{n-1}(x) (x - \alpha) d\psi(x) = 0;$$

while on the other hand by (13),

$$\int_a^b x^n U_{n-1}(x)(x-\alpha) d\psi(x) = \int_a^b x^n \frac{\phi_n(x)}{a_n(x-\alpha)} (x-\alpha) d\psi(x) = \frac{1}{a_n^2} \neq 0.$$

(ii) Here (43) becomes

$$(45) \quad U_{n+1}(x) = (x^2 + \bar{b}_{n+1}x + \bar{c}_{n+1})U_{n-1}(x) - l_{n+1}U_{n-2}(x).$$

We find as above (see (20)):

$$\begin{aligned} l_{n+1} &= \frac{\int_a^b x^n U_{n-1}(x) d\psi(x)}{\int_a^b x^{n-2} U_{n-2}(x) d\psi(x)} = -\frac{1}{a_n^2} \frac{a_{n-1}a_{n-2}\phi_{n-2}(\alpha)}{\phi_{n-1}(\alpha)}, \\ \bar{b}_{n+1} &= \frac{\int_a^b x^2 U_{n-1}^2(x) d\psi(x)}{\int_a^b U_{n-1}^2(x) d\psi(x)} = \bar{s}_{n-1} - S_{n+1} = \bar{s}_{n-1} - \bar{s}_{n+1}, \\ \bar{s}_{n+1} &= S_{n+1}. \end{aligned}$$

Multiplying (45) by $x^n(x-\alpha)d\psi(x)$, integrating, and making use of (10, 13, 2), we get further:

$$\begin{aligned} \left[S_{n+1}S_{n+2} - \frac{a_{n+2,n}}{a_{n+2}} \right] - \left[S_{n-1}S_n - \frac{a_{n,n-2}}{a_n} \right] \\ + \frac{a_{n-2}}{a_{n-1}} \frac{\phi_{n-2}(\alpha)}{\phi_{n-1}(\alpha)} + (\bar{s}_{n-1} - S_{n+1})S_{n+1} + \bar{c}_{n+1} = 0. \end{aligned}$$

Furthermore, using (12), we get

$$\begin{aligned} \bar{s}_{n-1} &= S_{n-1} - \frac{a_{n-2}\phi_{n-2}(\alpha)}{a_{n-1}\phi_{n-1}(\alpha)}, \\ \bar{c}_{n+1} &= \left[S_{n-1}S_n - \frac{a_{n,n-2}}{a_n} \right] - \left[S_{n+1}S_{n+2} - \frac{a_{n+2,n}}{a_{n+2}} \right] \\ &\quad + \frac{a_{n-2}\phi_{n-2}(\alpha)}{a_{n-1}\phi_{n-1}(\alpha)} c_{n+1} + S_{n+1}(S_{n+1} - S_{n-1}). \end{aligned}$$

In the so-called "symmetric" case $(\psi(x) \equiv -\psi(-x); (a, b) \equiv (-h, h))$

$$S_n = c_n = 0 \quad (n \geq 1), \quad \frac{a_{n,n-2}}{a_n} = -\sum_{i=2}^n \lambda_i,$$

so that

$$\bar{b}_{n+1} = \bar{s}_{n-1} = -\sqrt{\lambda_n} \frac{\phi_{n-2}(\alpha)}{\phi_{n-1}(\alpha)}, \quad \bar{c}_{n+1} = -\lambda_{n+2} - \lambda_{n+1} \quad (\text{"symmetric" case}).$$

(iii) $\phi_n(\alpha) = \phi_{n-2}(\alpha) = 0$. Proceeding as before, we get:

$$\begin{aligned} l_{n+1} &= \frac{a_{n-2}^2}{a_n^2}; \quad \bar{b}_{n+1} = \bar{s}_{n-1} - \bar{s}_{n+1} = S_{n-1} - S_{n+1}; \\ \bar{c}_{n+1} &= \left[S_{n-1}S_n - \frac{a_{n,n-2}}{a_n} \right] - \left[S_{n+1}S_{n+2} - \frac{a_{n+2,n}}{a_{n+2}} \right] + S_{n+1}(S_{n+1} - S_{n-1}); \\ \bar{b}_{n+1} &= 0; \quad \bar{c}_{n+1} = -\lambda_{n+2} - \lambda_{n-1} \quad (\text{"symmetric" case}). \end{aligned}$$

If $\phi_n(\alpha) = 0$, $\phi_{n+2}(\alpha) \neq 0$, we write

$$U_{n+2}(x) = (x - k_{n+2})U_{n+1}(x) + P_n(x), \text{ and } \int_a^b P_n(x)G_{n-1}(x)(x - \alpha)d\psi(x) = 0.$$

Hence,

$$(46) \quad P_n(x) = -l_{n+2}U_{n-1}(x), \quad U_{n+2}(x) = (x - k_{n+2})U_{n+1}(x) - l_{n+2}U_{n-1}(x),$$

and as above

$$l_{n+2} = -\frac{a_n^2}{a_{n+2}^2} \frac{\phi_{n+2}(\alpha)}{\phi_{n+1}(\alpha)}, \quad k_{n+2} = c_{n+2} - \frac{a_{n+1}}{a_{n+2}} \frac{\phi_{n+1}(\alpha)}{\phi_{n+2}(\alpha)}.$$

In order to illustrate we take in Theorem VI $r(x) \equiv x - \alpha$ and assume $\phi_1(\alpha)\phi_2(\alpha) \cdots \phi_{k-1}(\alpha) \neq 0$, $\phi_k(\alpha) = 0$. Then the polynomials $q_i(x)$ ($i = 1, 2, \dots, k-1$) in (41) are all of the first degree, while $q_k(x)$ is of the second degree. Correspondingly, the recurrence relation (19) holds for ($n = 1, 2, \dots, k-1$); for $n = k+1$ its character changes as indicated under (ii), (iii). (For $n = k$, (19) does not exist).

Case (iii) is possible. This is evident in the symmetric case with $\alpha = 0$, for here $\phi_{2n-1}(\alpha) = 0$ ($n = 1, 2, \dots$) so that all the $q_i(x)$ in (41) are of degree 2. This shows that the upper bound for the degree of $q_i(x)$ as given in Theorem VI is actually attained in this case.

14. A minimum property of $U_n(x)$. Among all polynomials $G_n(x)$ such that $G_n(\alpha) = 1$ ($\phi_n(\alpha) \neq 0$), it is the polynomial $\frac{K_n(x, \alpha)}{K_n(\alpha)} = \frac{a_n \phi_n(\alpha)}{K_n(\alpha)} U_n(x)$ which minimizes the integral $\int_a^b G_n^2(x) d\psi(x)$, with the minimum $\frac{1}{K_n(\alpha; d\psi)}$.

The proof can be easily accomplished by using the methods of constrained extrema.

15. The case of two changes of sign. We wish to investigate the existence of a system of polynomials

$$\{V_n(x) \equiv x^n + \cdots\} \quad (n = 0, 1, \dots)$$

satisfying the condition of orthogonality

$$(47) \quad \int_a^b V_n(x)G_{n-1}(x)d\psi(x), \text{ with } \psi(x) \equiv \int_a^x (t - \alpha_1)(t - \alpha_2)d\psi(t), \\ (a < \alpha_1 < \alpha_2 < b).$$

We get as above (see § 2), writing

$$(x - \alpha_1)(x - \alpha_2)V_n(x) \equiv \sum_{i=0}^{n+2} A_i \phi_i(x):$$

$$(48) \quad \begin{cases} A_{n+1}\phi_{n+1}(\alpha_1) + A_n\phi_n(\alpha_1) = -\frac{\phi_{n+2}(\alpha_1)}{a_{n+2}} \\ A_{n+1}\phi_{n+1}(\alpha_2) + A_n\phi_n(\alpha_2) = -\frac{\phi_{n+2}(\alpha_2)}{a_{n+2}}. \end{cases}$$

The determinant of (48) is $\frac{a_{n+1}}{a_n} (\alpha_1 - \alpha_2) K_n(\alpha_1, \alpha_2)$.

The condition $K_n(\alpha_1, \alpha_2) \neq 0$ is thus seen to be sufficient for the existence of $V_n(x)$ satisfying (47). Furthermore, this condition insures the unique determination of $V_n(x)$ in the form

$$V_n(x) = \frac{a_n}{a_{n+1}a_{n+2}(\alpha_1 - \alpha_2)K_n(\alpha_1, \alpha_2)(x - \alpha_1)(x - \alpha_2)} \begin{vmatrix} \phi_{n+2}(x) & \phi_{n+1}(x) & \phi_n(x) \\ \phi_{n+2}(\alpha_1) & \phi_{n+1}(\alpha_1) & \phi_n(\alpha_1) \\ \phi_{n+2}(\alpha_2) & \phi_{n+1}(\alpha_2) & \phi_n(\alpha_2) \end{vmatrix}.$$

In particular, if $K_n(\alpha_1, \alpha_2) \neq 0$ and $\phi_{n+2}(\alpha_1) = \phi_{n+2}(\alpha_2) = 0$,

$$V_n(x) = \frac{\phi_{n+2}(x)}{a_{n+2}(x - \alpha_1)(x - \alpha_2)}.$$

Moreover, in this latter case (47) is replaced by

$$\int_a^b V_n(x) G_{n+1}(x) d\psi(x) = 0.$$

On the other hand, if $K_n(\alpha_1, \alpha_2) = 0$, the consistency of the system (48) requires the matrix

$$\begin{vmatrix} \phi_{n+2}(\alpha_1) & \phi_{n+1}(\alpha_1) & \phi_n(\alpha_1) \\ \phi_{n+2}(\alpha_2) & \phi_{n+1}(\alpha_2) & \phi_n(\alpha_2) \end{vmatrix}$$

to be of rank one. We proceed to show that this is possible. In the first place, we must have

$$\begin{vmatrix} \phi_{n+2}(\alpha_1) & \phi_{n+1}(\alpha_1) \\ \phi_{n+2}(\alpha_2) & \phi_{n+1}(\alpha_2) \end{vmatrix} = \frac{a_{n+2}}{a_{n+1}} (\alpha_1 - \alpha_2) K_{n+1}(\alpha_1, \alpha_2) = 0, \quad K_{n+1}(\alpha_1, \alpha_2) = 0,$$

which, combined with the assumed relation $K_n(\alpha_1, \alpha_2) = 0$, gives

$$\phi_{n+1}(\alpha_1)\phi_{n+1}(\alpha_2) = 0.$$

If we assume now that $\phi_{n+1}(\alpha_1) = 0$ (hence $\phi_n(\alpha_1) \neq 0$), then (8) shows that $\phi_{n+1}(\alpha_2) = 0$. Conversely, assuming $\phi_{n+1}(\alpha_1) = \phi_{n+1}(\alpha_2) = 0$, we get at once $K_n(\alpha_1, \alpha_2) = K_{n+1}(\alpha_1, \alpha_2) = 0$.

In the second place, consider the last determinant of order 2 of our matrix and use (5, 2):

$$\begin{vmatrix} \phi_{n+2}(\alpha_1) & \phi_n(\alpha_1) \\ \phi_{n+2}(\alpha_2) & \phi_n(\alpha_2) \end{vmatrix} = \begin{vmatrix} (\alpha_1 - c_{n+2}) \frac{a_{n+2}}{a_{n+1}} \phi_{n+1}(\alpha_1) & -\lambda_{n+2} \phi_n(\alpha_1) \phi_n(\alpha_1) \\ (\alpha_2 - c_{n+2}) \frac{a_{n+2}}{a_{n+1}} \phi_{n+1}(\alpha_2) & -\lambda_{n+2} \phi_n(\alpha_2) \phi_n(\alpha_2) \end{vmatrix} = 0.$$

Hence, if $\phi_{n+1}(\alpha_1) = \phi_{n+1}(\alpha_2) = 0$, the determination of $V_n(x)$ by means of (48) is no longer unique. In this case A_{n+1} in (48) may be assigned arbitrarily. These considerations lead to

THEOREM VII. *A necessary and sufficient condition that a uniquely determined polynomial, of a given degree n , $V_n(x) = x^n + \dots$, satisfying (47), exist, is: $K_n(\alpha_1, \alpha_2) \neq 0$. Moreover, if $K_n(\alpha_1, \alpha_2) \neq 0$ for all n ,¹⁰ there exists a uniquely determined sequence $\{V_n(x)\}$ ($n = 0, 1, \dots$) of such polynomials.*

The polynomial $V_n(x)$, may have two (but not more than two) imaginary or equal zeros, or two (but not more than two) zeros outside the interval (a, b) . To show this construct the polynomial $\Phi_n(x; (x^2 + r^2)d\psi) = x^n + \dots$, (r arbitrary real constant), orthogonal with respect to the monotonic characteristic function $\int_a^x (t^2 + r^2)d\psi(t)$, and write the orthogonality property in the form

$$(49) \quad \int_a^b \frac{\Phi_n(x; (x^2 + r^2)d\psi)(x^2 + r^2)}{(x - \alpha_1)(x - \alpha_2)} G_{n-1}(x)(x - \alpha_1)(x - \alpha_2)d\psi(x) = 0,$$

where (α_1, α_2) are zeros of $\Phi_n(x; (x^2 + r^2)d\psi)$. (49) shows the existence of a polynomial

$$(50) \quad V_n(x) \equiv \frac{\Phi_n(x; (x^2 + r^2)d\psi)(x^2 + r^2)}{(x - \alpha_1)(x - \alpha_2)} = x^n + \dots,$$

orthogonal with respect to the characteristic function

$$\psi(x) = \int_a^x (t - \alpha_1)(t - \alpha_2)d\psi(t).$$

Furthermore, we readily see that here the determination of $V_n(x)$ is unique since this is known to be true for $\Phi_n(x; (x^2 + r^2)d\psi)$. Hence, (50) shows that with such choice of α_1, α_2 , $V_n(x)$ has two imaginary zeros. In like manner we show the possibility of the existence of two equal zeros or of two zeros outside (a, b) .

The case of s ($s > 2$) changes of sign could be treated as above. Since, however, even for two changes of sign the most important properties of the zeros of orthogonal polynomials corresponding to monotonic characteristic functions no longer hold, the discussion of this case is omitted.

UNIVERSITY OF PENNSYLVANIA.

¹⁰ There is an infinity of such α_1, α_2 in any subinterval of (a, b) .

METABELIAN GROUPS AND PENCILS OF BILINEAR FORMS.

By H. R. BRAHANA.

Introduction. In a recent paper¹ it was shown that the problem of classification of metabelian groups of order p^{n+m} which contain a given abelian group of order p^n as a maximal invariant abelian subgroup and have commutator subgroups of order p^m is equivalent to the problem of classification of the matrices $x_1M_1 + x_2M_2 + \cdots + x_kM_k$ under projective transformations on the x 's and elementary transformations on the square matrices M_1, M_2, \cdots, M_k . The x 's and the elements of the M 's are of course numbers in a modular field as are also the coefficients of the transformations. The squareness of the matrices comes from the requirement that the commutator subgroup be of order p^m . The situation may then be discussed in terms of the invariant factors of the matrix $x_1M_1 + x_2M_2 + \cdots + x_kM_k$.

The argument of that paper still holds when the commutator subgroup is not of order p^m and the matrices M_i are not square. In this case however we are deprived of the use of a well-developed theory of invariant factors. So far as I know the question of the conjugacy of two matrices of the above type under transformation on the x 's and simultaneous transformations on rectangular M 's has not been considered. It is our purpose to consider the groups which give rise to such matrices in the simple case where $m = 4$ and $k = 2$ and to use the results to obtain normal forms for the matrices. It will be convenient to interpret the matrices M_i as matrices of bilinear forms in which case the matrix above, which we shall denote hereafter as $\lambda_1M_1 + \lambda_2M_2$, may be taken to represent a pencil of bilinear forms.

1. *The groups.* We consider groups $G = \{H, U\}$ where H is abelian, of order p^n , and type $1, 1, \cdots$ and U is an abelian group of order p^4 and type $1, 1, \cdots$ from the group of isomorphisms of H . We require that no operator of U , except identity, be permutable with every operator of H . We require further that G be metabelian which implies that its commutator subgroup is in its central, that every operator of U determines² a partition of n with greatest term equal to 2. Finally, we require that no operator of U

¹ "Metabelian groups of order p^{n+m} with commutator subgroups of order p^m ," *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 776-792.

² "On metabelian groups," *American Journal of Mathematics*, vol. 56 (1934), pp. 490-510.

determine a partition of n in which more than two terms are equal to 2. An operator U_i will be said to be of type I or type II according as the partition it determines contains one or two 2's. The group $\{H, U_i\}$ will be said to be of type I or type II depending on the type of U_i . In the groups which we shall consider U will contain only operators of types I and II, the identity excepted. Since the groups in which U contains only operators of type I were classified in the paper just referred to we shall suppose that U contains at least one operator of type II.

The central of G under these conditions is of order p^{n-2} and we may suppose that generators of H are chosen so that all but two are in the central. Let the two of the generators of H which are not in the central be denoted by s_1 and s_2 . Then if U_i is an operator of U , $\{H, U_i\}$ will have a commutator subgroup of order p or p^2 according as it is of type I or type II. The maximum order for the commutator subgroup of G is p^8 and occurs only if each of the operators of every set of four which generate U is of type II and the resulting eight commutators are independent. The commutator subgroup of G has an order at least p^2 since U contains at least one operator of type II. Since the commutator subgroup is characteristic we may separate the groups in question into classes according to the orders of their commutator subgroups and no two groups belonging to different classes can be simply isomorphic.

Let the order of the commutator subgroup be p^l . It is immediately obvious that there is but one group in the class corresponding to $l=8$, for a simple isomorphism is established between any two such groups by a proper naming of generators. These considerations give the following more general theorem:

(1.1) *There is but one metabelian group $G = \{H, U\}$ of order p^{n+m} with commutator subgroup of order p^{2m} , provided the operators of U are restricted to types I and II. In this group no operator of U is of type I.*

If two groups U and U' have different numbers of subgroups composed of operators of type I, there will exist no simple isomorphism between $\{H, U\}$ and $\{H, U'\}$ in which H corresponds to itself.³ Accordingly, for $l < 8$ we may consider the groups in sets determined by the number of subgroups of type I in U .

When $l=7$, U cannot contain more than one subgroup of type I. Otherwise two operators U_1 and U_2 both of type I could be selected as two of the

³ We shall postpone the question of simple isomorphisms between $\{H, U\}$ and $\{H, U'\}$ in which H does not correspond to itself.

four independent generators of U . The group $\{H, U_1, U_2\}$ would have a commutator subgroup of order p^2 . The commutator subgroup of $\{H, U_3, U_4\}$ could be of order at most p^4 and hence the commutator subgroup of G could be of order at most p^6 . If U contains one subgroup of type I, let us suppose it to be generated by U_1 and generators of H chosen so that U_1 is permutable with s_2 . Then $\{H, U_2, U_3, U_4\}$ must have a commutator subgroup of order p^6 and by (1.1) just one such group exists. Consequently there exists such a group and it is completely determined by the requirements that $l=7$ and that U contain but one subgroup of type I.⁴ If U contains no operator of type I there is also but one group. No matter how U_1, \dots, U_4 are chosen, so long as they generate U , the group $\{H, U_1, U_2\}$ will have a commutator subgroup of order p^3 or p^4 and in the former case $\{H, U_3, U_4\}$ will have a commutator subgroup of order p^4 . We may then assume that the commutator subgroup K' of $\{H, U_1, U_2\}$ is of order p^4 . Then at most one of the commutators arising from transformation of H by U_3 is in K' and if so the two commutators arising from U_4 are independent of K' . Hence we may choose an operator U_3 such that the commutator subgroup K'' of $\{H, U_1, U_2, U_3\}$ is of order p^6 . Then of the two commutators arising from transformation of H by U_4 at most one is in K'' . From the symmetry in s_1 and s_2 of the relations which generators of G just described satisfy it is clear that no restriction is introduced by assuming that the commutator of U_4 and s_1 is in K'' . Let us denote this commutator by s_k , and the commutator of U_4 and s_2 by s_9 . If s_k is not in K'' it is in $\{K'', s_9\}$. If then we replace s_1 by a proper combination of s_1 and s_2 we obtain a commutator s'_k which is in K'' . We may assume further that s_k is in the part of K'' which arises from transformation of s_2 by U_1, U_2 , and U_3 , for otherwise U_4 could be replaced by such a combination of U_1, U_2, U_3 , and U_4 that such would be the case. Therefore there exists in $\{U_1, U_2, U_3\}$ an operator U' whose commutator with s_2 is s_k . This operator may be taken to be U_1 , and consequently we may assume that the commutator of U_1 and s_2 is the same as that of U_4 and s_1 . There are therefore two groups for $l=7$ and they are distinguished by the numbers of subgroups of type I in U . These considerations also apply more generally to give the theorem:

(1.2) *If the operators of U are all of types I and II, there are two and only two groups of the type we are considering of order p^{n+m} with commutator subgroups of order p^{2m-1} . In one of them U contains one subgroup of type I, and in the other none.*

⁴ We use the shorter expression "subgroup of type I" in place of "subgroup composed of operators of type I, and the identity."

When $l=6$ the number of independent subgroups of type I in U can be at most two as may be seen from an argument similar to that used for $l=7$. We shall see that U may contain more than two subgroups of type I but that if so all such subgroups are contained in the group generated by two of them. The possibilities for the number of independent subgroups of type I are therefore 2, 1, and 0.

If U contains two independent subgroups of type I, let them be generated by U_3 and U_4 . Two possibilities arise: the subgroups of $\{s_1, s_2\}$ permutable with the respective operators U_3 and U_4 may be the same or they may be different. If they are the same every operator of $\{U_3, U_4\}$ is of type I; if they are different every operator of $\{U_3, U_4\}$ except those in $\{U_3\}$ and $\{U_4\}$ is of type II. In either case since the commutator subgroup of $\{H, U_1, U_2\}$ is of order at most p^4 , the commutator subgroup of $\{H, U_3, U_4\}$ must be of order p^2 . In either case every operator of U not in $\{U_3, U_4\}$ is of type II. Hence, in the one case U contains $1+p$ subgroups of type I, and in the other contains two subgroups of type I. The two groups are generated by operators $s_1, s_2, s_3, \dots, s_n, U_1, \dots, U_4$ which satisfy the following relations: when U contains $1+p$ subgroups of type I,

$$(1.3) \quad \begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_5, & U_3^{-1}s_1U_3 &= s_1s_7, & U_4^{-1}s_1U_4 &= s_1s_8, \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_2^{-1}s_2U_2 &= s_2s_6; \end{aligned}$$

when U contains two subgroups of type I,

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_5, & U_3^{-1}s_1U_3 &= s_1s_7, \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_3^{-1}s_2U_3 &= s_2s_6, & U_4^{-1}s_2U_4 &= s_2s_8. \end{aligned}$$

If U contains one subgroup of type I let it be generated by U_4 and let U_4 be permutable with s_2 . The commutator subgroup of $\{H, U_1, U_2, U_3\}$ must be of order p^5 or p^6 , and $\{U_1, U_2, U_3\}$ can contain no operator of type I. By theorems (1.1) and (1.2) there is but one group in each case. In the former case the group is completely determined, since the commutators of $\{H, U_4\}$ and $\{H, U_1, U_2, U_3\}$ are independent. A set of generating relations is obtained by adding $U_3^{-1}s_2U_3 = s_2s_3$ to (1.3) above. In the second case we may obtain such a group by adding the relation $U_3^{-1}s_2U_3 = s_2s_k$ to (1.3). It then follows that s_8 is in the group $\{s_3, s_4, s_5, s_6, s_7, s_k\}$. The group $\{s_3, s_5, s_7, s_8\}$ must be of order p^4 , otherwise a proper combination U' of U_1, U_2 , and U_3 would transform s_1 into s_1s_8 and $U'U_4^{-1}$ would be of type I contrary to the assumption that U contained no operator of type I besides U_4 . At least two of the operators s_4, s_6 , and s_k are independent of $\{s_3, s_5, s_7, s_8\}$; we may then suppose that s_8 is expressible in terms of s_3, s_4, s_5, s_6, s_7 , and s_k .

If the group which we are considering is distinct from the one just previously described, we must expect at least two possibilities to appear, for if in the former group we replace U_1 by U_1U_4 the commutator subgroup of $\{H, U_1, U_2, U_3\}$ is of order p^6 , being generated by $s_3s_8, s_4, s_5, s_6, s_7$, and s_8 . There are then at least these two possibilities in the present case: (a) s_8 is in the group $\{s_3, s_4, s_5, s_6, s_7, s_k\}$ and not in $\{s_4, s_6, s_k\}$, or (b) s_8 is in $\{s_4, s_6, s_k\}$. In case (b) we may assume that $s_k = s_8$, for every operator of $\{s_4, s_6, s_k\}$ is a commutator, and any set of three independent operators of $\{U_1, U_2, U_3\}$ generate it. The operator U_4 is a unique operator in U , it defines s_2 to within a power of a single operator of $\{s_1, s_2\}$, and s_2 in turn defines $\{s_4, s_6, s_k\}$. Hence, this group is distinct from the one previously defined, in which s_8 is not in $\{s_4, s_6, s_k\}$ and as we have seen may lead to case (a). We complete this by showing that case (a) leads to a single group, consequently one in which U_1, U_2, U_3 can be chosen so that $\{H, U_1, U_2, U_3\}$ has a commutator subgroup of order p^5 . If s_8 is in $\{s_3, \dots, s_7, s_k\}$ but not in $\{s_4, s_6, s_k\}$, then there exists in $\{U_1, U_2, U_3\}$ an operator U'_1 and in $\{s_1, s_2\}$ an operator s'_1 such that the commutator of U'_1 and s'_1 is s_ks_8 . This operator s'_1 is obviously not a power of s_2 . The operator U'_1 is not U_3 for in that case $U_3U_4^{-1}$ would give the same commutator with s'_1 as with s_2 and hence would be of type I. Consequently, if we replace U_1 by $U'_1U_4^{-1}$ and s_1 by s'_1 , generators of the group satisfy the relations found for the first group with one subgroup of type I.

The two groups each containing one subgroup of type I just described have been distinguished by the order of the group $\{s_4, s_6, s_k, s_8\}$ which in one case is p^3 and in the other p^4 . They may also be distinguished from each other by the non-abelian subgroups which they contain. In one there exists at least one subgroup $\{H, U_1, U_2, U_3\}$ of order p^{n+3} with commutator subgroup of order p^5 which contains no subgroup of type I; in the other every subgroup of order p^{n+3} with commutator subgroup of order p^5 contains a subgroup of type I. Thus there is no simple isomorphism between the two groups in which H corresponds to H .⁵

Now suppose that U contains no operator of type I. Two possibilities are immediately evident: (a) U contains two operators U_1 and U_2 such that $\{H, U_1, U_2\}$ has a commutator subgroup of order p^2 ; or (b) U contains no

⁵ We beg leave to point out that all the operators of type II in the group of isomorphisms of H are conjugate, as are all the operators of type I. The two groups U and U' are abelian, of order p^4 , and type 1, 1, . . . and every operator of one can be transformed into many operators of the other by operators which transform H into itself. U may be transformed into U' in many ways. U may not, however, be transformed into U' by any operator which leaves H invariant.

such group. The condition (a) completely determines the group, for the commutator subgroup of $\{H, U_3, U_4\}$ must then be of order p^4 and must be independent of the commutator subgroup of $\{H, U_1, U_2\}$. In any case, no matter how s_1 and s_2 are chosen the commutator subgroups H_1 and H_2 arising from transformation of s_1 and s_2 respectively by U are of order p^4 , since U contains no operator of type I. Since $\{H_1, H_2\}$ is of order p^6 , these two groups have a subgroup of order p^2 in common. This subgroup determines two operators U_1 and U_2 such that the commutator subgroup arising from transformation of s_1 by $\{U_1, U_2\}$ is the subgroup common to H_1 and H_2 ; it determines two other operators V_1 and V_2 such that the commutator subgroup arising from transformation of s_2 by $\{V_1, V_2\}$ is the same group. In case (a) the group $\{U_1, U_2, V_1, V_2\}$ is of order p^2 . In case (b) this group is of order at least p^3 . In case it is of order p^4 , the four commutators arising from transformation of s_1 by V_1 and V_2 and of s_2 by U_1 and U_2 must be independent of the commutator subgroup common to H_1 and H_2 . The group is therefore completely determined. If the group $\{U_1, U_2, V_1, V_2\}$ is of order p^3 , it is generated by three operators U_1, U_2 , and U_3 and its commutator subgroup is of order p^4 . Then the commutator subgroup of $\{H, U_4\}$ must be independent of this group of order p^4 , and since there is but one group of order p^{n+3} with commutator subgroup of order p^4 and no subgroups of type I⁶ it is completely determined. It is necessary to determine whether or not these last two groups are distinct. This last group contains a subgroup of order p^{n+3} with commutator subgroup of order p^4 and the preceding group does not.

Normal forms for generating relations of the three groups with commutator subgroups of order p^6 and no subgroups of type I are:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_4r, & U_3^{-1}s_1U_3 &= s_1s_5, & U_4^{-1}s_1U_4 &= s_1s_7, \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_2^{-1}s_2U_2 &= s_2s_3, & U_3^{-1}s_2U_3 &= s_2s_6, & U_4^{-1}s_2U_4 &= s_2s_8, \end{aligned}$$

where r is any not-square. This group contains a subgroup of order p^{n+2} with commutator subgroup of order p^2 .

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_5, & U_3^{-1}s_1U_3 &= s_1s_6, & U_4^{-1}s_1U_4 &= s_1s_7, \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_2^{-1}s_2U_2 &= s_2s_3, & U_3^{-1}s_2U_3 &= s_2s_5, & U_4^{-1}s_2U_4 &= s_2s_8. \end{aligned}$$

This group contains no subgroup of order p^{n+2} with commutator subgroup of order p^2 , but contains one of order p^{n+3} with commutator subgroup of order p^4 .

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_5, & U_3^{-1}s_1U_3 &= s_1s_7, & U_4^{-1}s_1U_4 &= s_1s_8, \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_2^{-1}s_2U_2 &= s_2s_6, & U_3^{-1}s_2U_3 &= s_2s_3, & U_4^{-1}s_2U_4 &= s_2s_5. \end{aligned}$$

⁶ "On metabelian groups," *loc. cit.*, p. 510.

This group contains no subgroup of order p^{n+3} with commutator subgroup of order p^4 .

As in the cases of $l = 8$ and $l = 7$, an obvious rewording of the argument above gives the more general theorem:

(1.4) *If the operators of U are all of types I and II there are seven and only seven groups of the type we are considering of order p^{n+m} with commutator subgroups of order p^{2m-2} . They may be characterized by their non-abelian subgroups of orders p^{n+1} , p^{n+2} , and p^{n+3} .*

When $l = 5$, the number of independent subgroups of type I in U cannot be greater than 3, and may be 3, 2, 1, or 0. If there are 3 independent subgroups of type I in U we may take them to be generated by U_2 , U_3 , and U_4 . The operator U_1 must be of type II. The commutator subgroup of $\{H, U_2, U_3, U_4\}$ must be of order p^3 and must have no operator in common with the commutator subgroup of $\{H, U_1\}$. The groups will then be characterized by the properties of $\{H, U_2, U_3, U_4\}$. Each of the operators U_2 , U_3 , and U_4 is permutable with a subgroup of order p of $\{s_1, s_2\}$. These three subgroups may be the same; two of them may coincide; or all three may be distinct. If the three subgroups coincide let us suppose the subgroup to be generated by s_2 . Then generators of G will satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_5, & U_3^{-1}s_1U_3 &= s_1s_6, & U_4^{-1}s_1U_4 &= s_1s_7, \\ U_1^{-1}s_2U_1 &= s_2s_4. \end{aligned}$$

In this case every operator of $\{U_2, U_3, U_4\}$ will be of type I, and G will contain $1 + p + p^2$ subgroups of type I.

In the second case the two subgroups of $\{s_1, s_2\}$ permutable with operators of type I of U may be taken to be generated by s_1 and s_2 . In this case generators of G will satisfy relations the same as those above except that $U_4^{-1}s_1U_4 = s_1s_7$ is replaced by $U_4^{-1}s_2U_4 = s_2s_7$. The only operators of type I in $\{U_2, U_3, U_4\}$ are the operators of $\{U_2, U_3\}$ and powers of U_4 . G therefore contains $2 + p$ subgroups of type I.

In the third case we may suppose that two of the groups are generated by s_1 and s_2 respectively. Generators of G will satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_5, & U_3^{-1}s_1U_3 &= s_1s_6, \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_3^{-1}s_2U_3 &= s_2s_6, & U_4^{-1}s_2U_4 &= s_2s_7. \end{aligned}$$

It is obvious that the group $\{U_2, U_3, U_4\}$ contains but three subgroups of type I, and that G likewise contains but three subgroups of type I.

When U contains but two independent subgroups of type I the operators of $\{s_1, s_2\}$ permutable respectively with these two subgroups of U may constitute one or two subgroups of order p . If they constitute one subgroup of order p , let it be generated by s_2 . Let U_3 and U_4 generate the respective subgroups of type I. Then $\{U_1, U_2\}$ can contain only operators of type II. The commutator subgroup of $\{H, U_1, U_2\}$ is of order p^3 or p^4 . There is but one group for each of the orders p^3 and p^4 having the required properties. If the order of the commutator subgroup of $\{H, U_1, U_2\}$ is p^3 the group G is completely determined. Its generators satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_5, & U_3^{-1}s_1U_3 &= s_1s_6, & U_4^{-1}s_1U_4 &= s_1s_7, \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_2^{-1}s_2U_2 &= s_2s_3. \end{aligned}$$

If the commutator subgroup of $\{H, U_1, U_2\}$ is of order p^4 , then it must have a subgroup of order p in common with the commutator subgroup of $\{H, U_3, U_4\}$. This common subgroup may be taken to be in the group of commutators arising from transformation of s_2 by U_1 and U_2 , for otherwise U'_1 and U'_2 could be chosen so that the commutator subgroup of $\{H, U'_1, U'_2\}$ would be of order p^3 . The group G in this case is also completely determined. Its generators satisfy relations obtained from these above by changing the commutator of U_2 and s_2 from s_3 to s_6 . The two groups are obviously distinct; each contains $1 + p$ subgroups of type I. They may be distinguished by the fact that the first contains a subgroup of order p^{n+2} with commutator subgroup of order p^3 and no subgroup of type I, whereas in the second every subgroup of order p^{n+2} with commutator subgroup of order p^3 contains subgroups of type I.

If the subgroups of $\{s_1, s_2\}$ permutable respectively with U_3 and U_4 are distinct let them be generated by s_1 and s_2 . The commutator subgroups of $\{H, U_3\}$ and $\{H, U_4\}$ may coincide, in which case the commutator subgroup of $\{H, U_1, U_2\}$ must be of order p^4 and independent of it. Generators of G therefore satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_5, & U_3^{-1}s_1U_3 &= s_1s_7, \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_2^{-1}s_2U_2 &= s_2s_6, & U_4^{-1}s_2U_4 &= s_2s_7. \end{aligned}$$

In this case G contains $1 + p$ subgroups of type I. The group is obviously distinct from the three preceding ones.

If the commutator subgroup of $\{H, U_3, U_4\}$ is of order p^2 , there are but two subgroups of type I in $\{U_3, U_4\}$ and therefore but two in U . The commutator subgroup of $\{H, U_1, U_2\}$ is of order p^3 or p^4 . In the first case generators of G satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_5, & U_3^{-1}s_1U_3 &= s_1s_6, \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_2^{-1}s_2U_2 &= s_2s_3, & U_4^{-1}s_2U_4 &= s_2s_7. \end{aligned}$$

In the other case generators of G satisfy the above relations with s_2 replaced by s_k in the transform of s_2 by U_2 . The operator s_k cannot be in the group $\{s_3, s_4, s_5\}$, but must be in the group $\{s_3, s_4, s_5, s_6, s_7\}$. We may assume that the expression for s_k in terms of these operators does not contain s_4 or s_7 . It is therefore in the group $\{s_3, s_5, s_6\}$. There is an operator in the group $\{U_1, U_2, U_3\}$ whose commutator with s_1 is s_k , and this operator is not U_2 . If it is not U_3 , it may serve in place of U_1 and generators of G satisfy the relations above. If it is U_3 , we have a new group whose generators satisfy the above relations with s_6 for the commutator of U_2 and s_2 . This last group contains no subgroup of order p^{n+2} with commutator subgroup of order p^3 except those which contain subgroups of type I; the former group does contain such subgroups.

When U contains but one subgroup of type I let it be generated by U_4 and let the subgroup of $\{s_1, s_2\}$ permutable with U_4 be generated by s_2 . Let $U_4^{-1}s_1U_4 = s_1s_7$. The commutator subgroup of $\{H, U_1, U_2, U_3\}$ is then of order p^4 or p^5 . If it is of order p^4 it does not contain s_7 . The two groups of commutators, H_1 and H_2 , obtained by transforming s_1 and s_2 respectively by $\{U_1, U_2, U_3\}$ have a subgroup of order p^2 in common. This subgroup may be taken to be $\{s_3, s_4\}$ and it defines two operators U_1 and U_2 which give commutators s_3 and s_4 with s_1 and two operators V_1 and V_2 which give commutators s_3 and s_4 with s_2 . The group $\{U_1, U_2, V_1, V_2\}$ is of order p^2 or p^3 . We have the following two groups:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_4, & U_3^{-1}s_1U_3 &= s_1s_5, & U_4^{-1}s_1U_4 &= s_1s_7, \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_2^{-1}s_2U_2 &= s_2s_3, & U_3^{-1}s_2U_3 &= s_2s_6. \end{aligned}$$

This group contains a subgroup of order p^{n+2} with commutator subgroup of order p^2 .

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_4, & U_3^{-1}s_1U_3 &= s_1s_6, & U_4^{-1}s_1U_4 &= s_1s_7, \\ U_1^{-1}s_2U_1 &= s_2s_5, & U_2^{-1}s_2U_2 &= s_2s_3, & U_3^{-1}s_2U_3 &= s_2s_4. \end{aligned}$$

This group contains no subgroup of order p^{n+2} with commutator subgroup of order p^2 . These are the only two groups when U_1, U_2 , and U_3 can be selected so that s_7 is not in the commutator subgroup of $\{H, U_1, U_2, U_3\}$.

If G contains no subgroup $\{H, U_1, U_2, U_3\}$ with commutator subgroup of order p^4 , we may assume that H_1 and H_2 have a subgroup of order p in common. Generators of G will satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_5, & U_3^{-1}s_1U_3 &= s_1s_6, & U_4^{-1}s_1U_4 &= s_1s_7, \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_2^{-1}s_2U_2 &= s_2s_3, & U_3^{-1}s_2U_3 &= s_2s_8. \end{aligned}$$

The operator s_k is in the group $\{s_3, \dots, s_7\}$ and it is not in the group $\{s_3, s_4, s_5, s_6\}$. In the expression for s_k neither s_3 nor s_4 need appear. It may therefore be assumed to be in the group $\{s_5, s_6, s_7\}$, but not in the group $\{s_5, s_6\}$. There exists in $\{U_2, U_3, U_4\}$ an operator U'_2 whose commutator with s_1 is s_k . This operator is not U_3 and may be used in place of U_2 if it is not U_4 . If it is not U_4 then H_1 and H_2 have the group $\{s_3, s_k\}$ of order p^2 in common, which brings us back to the group previously described. We may then assume in this case that $s_k = s_7$. There are thus three distinct groups with commutator subgroup of order p^5 , each containing one subgroup of type I. They are distinguished by means of their subgroups of orders p^{n+2} and p^{n+3} .

When U contains no operator of type I the two groups H_1 and H_2 are each of order p^4 and consequently have a subgroup of order p^3 in common. This subgroup determines three operators U_1, U_2 , and U_3 such that the commutator subgroup arising from transformation of s_1 by $\{U_1, U_2, U_3\}$ is the common subgroup of H_1 and H_2 . There are likewise three operators V_1, V_2 , and V_3 determined by s_2 and the common subgroup. The order of $\{U_1, \dots, V_3\}$ is either p^3 or p^4 . If it is of order p^3 it is generated by U_1, U_2 , and U_3 . There is but one⁷ such group $\{H, U_1, U_2, U_3\}$ with commutator subgroup of order p^3 , and since its commutator subgroup must be independent of that of $\{H, U_4\}$ the group G is completely determined. Its generators satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_5, & U_3^{-1}s_1U_3 &= s_1s_3^{\alpha}s_4^{\beta}, & U_4^{-1}s_1U_4 &= s_1s_6 \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_2^{-1}s_2U_2 &= s_2s_3, & U_3^{-1}s_2U_3 &= s_2s_5, & U_4^{-1}s_2U_4 &= s_2s_7, \end{aligned}$$

where $x^3 - \alpha x + \beta \equiv 0$ is irreducible, mod p .

If the order of $\{U_1, \dots, V_3\}$ is p^4 two possibilities arise according as $\{U_1, \dots, V_3\}$ does or does not contain a subgroup such that $\{H, U'_1, U'_2\}$ has a commutator subgroup of order p^2 . In the first case generators of G satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_4^r, & U_3^{-1}s_1U_3 &= s_1s_6, & U_4^{-1}s_1U_4 &= s_1s_7, \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_2^{-1}s_2U_2 &= s_2s_3, & U_3^{-1}s_2U_3 &= s_2s_6, & U_4^{-1}s_2U_4 &= s_2s_5. \end{aligned}$$

In the second case generators of G satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_5, & U_3^{-1}s_1U_3 &= s_1s_6, & U_4^{-1}s_1U_4 &= s_1s_7, \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_2^{-1}s_2U_2 &= s_2s_3, & U_3^{-1}s_2U_3 &= s_2s_5, & U_4^{-1}s_2U_4 &= s_2s_6. \end{aligned}$$

There are thus thirteen groups for $l = 5$.

⁷ Cf. "On metabelian groups," *loc. cit.*

In the case where $l = 4$ we shall use the results of the paper ⁸ on metabelian groups of order p^{n+m} with commutator subgroups of order p^m . Two matrices M and N are used to describe the commutator structure of $\{s_1, s_3, s_4, \dots, s_n, U_1, \dots, U_4\}$ and $\{s_2, s_3, \dots, s_n, U_1, \dots, U_4\}$ respectively. If the commutator subgroup of G is $\{s_3, s_4, s_5, s_6\}$, the element m_{ij} of M is the exponent of s_{j+2} in the commutator of U_i and s_1 ; and the element n_{ij} of N is the exponent of s_{j+2} in the commutator of U_i and s_2 . The groups G may then be classified according to the classification of the matrices $M + \lambda N$ under elementary transformations on the matrices M and N simultaneously and projective transformations on λ , both sorts of transformation having coefficients in the modular field, mod p .

Let us consider the determinant $f(\lambda) = |M + \lambda N|$, and suppose that $f(\lambda)$ is the product of four linear factors. If λ_1 is a root of $f(\lambda) \equiv 0$ then $s_1 s_2^{\lambda_1}$ is permutable with some operator of U . If all four roots of $f(\lambda) \equiv 0$ were the same we could take λ_1 to be zero and every element of M would be zero. This would imply that s_1 was permutable with every operator of U in which case U would contain only operators of type I. Since we assume U to contain at least one operator of type II, $f(\lambda)$ cannot be the fourth power of a linear expression in λ . We may then suppose that at least two of the linear factors of $f(\lambda)$ are distinct and that their zeros are 0 and ∞ . The determinants of M and N are then both zero. If we consider first the case where $f(\lambda)$ has just two distinct zeros we have the two possibilities (a) $f(\lambda) = \lambda^3$ and (b) $f(\lambda) = \lambda^2$. In the first case three of the U 's are permutable with s_1 and one with s_2 so that G contains $2 + p + p^2$ subgroups of type I. In the second case two U 's are permutable with each so that G contains $2(1 + p)$ subgroups of type I.

When $f(\lambda)$ is the product of four linear factors three of which are distinct we may suppose its zeros to be 0, 1, and ∞ with 0 counted twice. Then two U 's are permutable with s_1 . In this case G contains $3 + p$ subgroups of type I.

When $f(\lambda)$ has four distinct zeros they may be taken to be 0, 1, ∞ , and ρ , where ρ is the cross-ratio of the four taken in some arbitrary order. There are as many such groups as there are projectively distinct unordered sets of four points on the finite line, mod p . The cases where $\rho = 0, 1$, and ∞ give the group described just above. The values 2, -1 , $1/2$ give a single group, and the two primitive cube roots of -1 , when they exist in the modular field, give a single group. The rest of the numbers in the modular field go in sets of six to determine a single group. There are therefore $(p + 1)/6$ or

⁸ *Loc. cit.*

$(p+5)/6$, according as p is of the form $6k-1$ or $6k+1$, groups which are distinct from each other and from the groups considered above. Each has four subgroups of type I.

When $f(\lambda)$ does not have four linear factors it has at most two. Let us consider the case where it has just two linear factors. If they are the same, we may suppose that $f(\lambda) = q(\lambda) \cdot \lambda^2$, where $q(\lambda)$ is an irreducible quadratic. All such quartics are conjugate under the projective group on λ , and hence there is but one such group. The group G contains $1+p$ subgroups of type I, and it contains a subgroup of order p^{n+2} with commutator subgroup of order p^2 and no operators of type I; this last subgroup corresponds to the irreducible quadratic.

If the two linear factors of $f(\lambda)$ are distinct we have $f(\lambda) = q(\lambda) \cdot \lambda$. U contains one subgroup permutable with s_1 and one permutable with s_2 . G contains therefore two subgroups of type I and a subgroup of order p^{n+2} with commutator subgroup of order p^2 and no operator of type I. The number of distinct such groups is the number of conjugate sets of polynomials $q(\lambda)(\lambda-\lambda_1)(\lambda-\lambda_2)$ under the "rational" projective group on λ . The irreducible quadratic may be transformed into any particular quadratic and there is then a group of order $2(p+1)$ which leaves it fixed. There exists an operator of order two which leaves $q(\lambda)$ and $f(\lambda)$ fixed, viz., the one which interchanges the zeros of $q(\lambda)$ and also interchanges λ_1 and λ_2 . Consequently there are $p+1$ quadratics $(\lambda-\lambda_1)(\lambda-\lambda_2)$ such that $q(\lambda)(\lambda-\lambda_1)(\lambda-\lambda_2)$ belong to the same conjugate set unless λ_1, λ_2 , and the zeros of $q(\lambda)$ constitute a harmonic set in which case there are $(p+1)/2$ in the conjugate set. Of the first kind there are therefore $(p-1)/2$ conjugate sets and of the second kind one. There are in all $(p+1)/2$ groups of this kind.

When $f(\lambda)$ has but one linear factor the other factor is an irreducible cubic. G has one subgroup of type one, and a subgroup of order p^{n+3} with commutator subgroup of order p^3 and no subgroup of type I. To determine the number of such groups, let us write $f(\lambda) = c(\lambda)(\lambda-\lambda_1)$, where $c(\lambda)$ is the irreducible cubic. The "rational" projective group contains an operator of order three which transforms a root λ_0 of $c(\lambda) \equiv 0$ into its p -th power.⁹ This group associates three numbers λ_1, λ_2 , and λ_3 with any one of them, so that the cross-ratio of $\lambda_0, \lambda_0^p, \lambda_0^{p^2}, \lambda$, equals the cross-ratio of $\lambda_0^p, \lambda_0^{p^2}, \lambda_0, \lambda_2$ equals the cross-ratio of $\lambda_0^{p^2}, \lambda_0, \lambda_0^p, \lambda_3$. There are therefore $(p+5)/3$ or $(p+1)/3$ conjugate sets of such quartics depending on whether p is of the form $6k+1$ or $6k-1$.¹⁰

⁹ Cf. "On cubic congruences," *Bulletin of the American Mathematical Society*, vol. 39 (1933), pp. 962-969.

¹⁰ The cubic $c(\lambda)$ appears as a quartic with one root infinite. And when

There remain the groups G with no subgroups of type I. There are $p + 2$ such groups: There are thus $2p + 7$ or $2p + 9$ groups with commutator subgroups of order p^4 according as p is of the form $6k - 1$ or $6k + 1$.

Let us now suppose that $l = 3$. Then however generators of $\{s_1, s_2\}$ are selected, H_1 and H_2 are of orders not greater than p^3 . Consequently, U contains at least two subgroups of type I. Let us suppose that U contains four independent operators of type I. The operators of $\{s_1, s_2\}$ permutable with them must constitute at least two subgroups of order p . If they constitute two subgroups, then two possibilities arise: (a) three of the four independent U 's are permutable with s_1 and the other with s_2 , or (b) two U 's are permutable with s_1 and two with s_2 . In case (a) generators of G satisfy the relations:

$$U_1^{-1}s_1U_1 = s_1s_3, \quad U_2^{-1}s_1U_2 = s_1s_4, \quad U_3^{-1}s_1U_3 = s_1s_5, \\ U_4^{-1}s_2U_4 = s_2s_5.$$

In this case the group $\{U_1, U_2, U_3\}$ contains only operators of type I, as does $\{U_3, U_4\}$. There are therefore $1 + p + p^2$ such subgroups in the first, and $1 + p$ in the second; only the subgroup generated by U_3 is in both. Hence G contains $1 + 2p + p^2$ subgroups of type I. In case (b) generators of G satisfy the relations:

$$U_1^{-1}s_1U_1 = s_1s_3, \quad U_2^{-1}s_1U_2 = s_2s_4, \\ U_3^{-1}s_2U_3 = s_2s_4, \quad U_4^{-1}s_2U_4 = s_2s_5.$$

In this case $\{U_1, U_2\}$, $\{U_2, U_3\}$, and $\{U_3, U_4\}$ each contain $1 + p$ subgroups of type I of which two, generated by U_2 and U_3 respectively, are counted twice each. G therefore contains $1 + 3p$ subgroups of type I.

When the four independent U 's of type I are permutable with three subgroups of order p of $\{s_1, s_2\}$, we may suppose two of them to be permutable with s_1 , one with s_2 , and the other with $s_1s_2^{-1}$. The group H_1 must be of order p^3 , otherwise U_3 could be selected so that it was permutable with s_1 . Generators of G satisfy the relations:

$$U_1^{-1}s_1U_1 = s_1s_3, \quad U_2^{-1}s_1U_2 = s_1s_4, \quad U_3^{-1}s_1U_3 = s_1s_5, \\ U_3^{-1}s_2U_3 = s_2s_5, \quad U_4^{-1}s_2U_4 = s_2s_3.$$

G contains $1 + 3p$ subgroups of type I.

$p = 6k + 1$, $c(\lambda)$ may be taken to be $\lambda^3 + \alpha$, where $-\alpha$ is not a cube; the group leaving $c(\lambda)$ fixed is generated by $\lambda' = \rho\lambda$, where ρ is a primitive cube root of unity. This leaves $c(\lambda)$ and $c(\lambda) \cdot \lambda$ fixed; the remaining $6k$ quartics are separated into $2k$ sets of 3 each. We have thus the $(p + 5)/3$ sets.

When the four independent U 's are permutable with different subgroups of $\{s_1, s_2\}$ we may suppose U_1 permutable with s_2 , U_2 with $s_1s_2^{-1}$, U_3 with $s_1s_2^{-r}$, and U_4 with s_1 . Generators of G satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_4, & U_3^{-1}s_1U_3 &= s_1s_5^r, \\ U_2^{-1}s_2U_2 &= s_2s_4, & U_3^{-1}s_2U_3 &= s_2s_5, & U_4^{-1}s_2U_4 &= s_2s_3. \end{aligned}$$

The groups $\{U_1, U_4\}$, $\{U_1U_4, U_2\}$, $\{U_1^rU_4, U_3\}$ each contain $1 + p$ subgroups of type I, of which those generated by U_1U_4 and $U_1^rU_4$ are counted twice. G has therefore $1 + 3p$ subgroups of type I.

Of the groups just described the last contains four independent U 's permutable with three subgroups of $\{s_1, s_2\}$, viz., U_1 permutable with s_2 , U_2 and U_1U_4 permutable with $s'_1 = s_1s_2^{-1}$, and U_3 permutable with $s_1^{-1}s_2^r$. Hence this group is simply isomorphic with the preceding one. The preceding one itself contains four independent U 's permutable with two subgroups of $\{s_1, s_2\}$, viz., U_1 and U_2 permutable with s_2 , and U_3 and U_1U_4 permutable with $s'_1 = s_1s_2^{-1}$. This is therefore simply isomorphic with the one which precedes it.

We now suppose that U contains three independent operators of type I. They cannot all be permutable with the same subgroup of $\{s_1, s_2\}$. Suppose first they are permutable with two subgroups. Then generators of G satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_4, & U_3^{-1}s_1U_3 &= s_1s_5, \\ U_1^{-1}s_2U_1 &= s_2s_j, & & & U_4^{-1}s_2U_4 &= s_2s_k. \end{aligned}$$

The group H_1 must be of order p^3 if $\{U_1, U_2, U_3\}$ contains no operators of type I except those in $\{U_2, U_3\}$, and consequently the commutators in the first row may be taken to be s_3, s_4 , and s_5 . Now s_j is in $\{s_3, s_4, s_5\}$ and consequently $\{U_1, U_2, U_3\}$ does contain operators of type I not in $\{U_2, U_3\}$. Hence the supposition that there are not more than three independent U 's of type I and that they are permutable with but two subgroups of $\{s_1, s_2\}$ leads to a contradiction.

We suppose then that the three U 's are permutable with three subgroups of $\{s_1, s_2\}$. Generators of G satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_j, & U_2^{-1}s_1U_2 &= s_1s_3, & U_3^{-1}s_1U_3 &= s_1s_4, \\ U_1^{-1}s_2U_1 &= s_2s_k, & & & U_3^{-1}s_2U_3 &= s_2s_4, & U_4^{-1}s_2U_4 &= s_2s_5, \end{aligned}$$

For if s_3 and s_5 were the same, generators of $\{U_2, U_3, U_4\}$ could be selected to give the case above. Now s_j is in $\{s_3, s_4, s_5\}$ and since U_4 may be replaced by any power of itself s_j may be taken to be s_5 . Then s_k may be supposed

to be in the group $\{s_3, s_4\}$. Since U_2 may be replaced by any power of itself and U_3 likewise, we may assume that s_k is s_3, s_4 , or s_3s_4 . If $s_k = s_3$ then the operator $U_1U_2U_3$ is of type I, and U contains four independent operators of type I. If $s_k = s_4$, then $U_1U_3^{-1}$ is of type I. The only possibility is that $s_k = s_3s_4$. The group G exists, it contains but three subgroups of type I, and is therefore distinct from any of those which precede.

When U contains just two independent operators of type I they are permutable with different subgroups of $\{s_1, s_2\}$. Two possibilities arise: (a) U contains $1 + p$ subgroups of type I, or (b) U contains two subgroups of type I.

Let the two independent U 's be U_3 and U_4 . In case (a) the commutator subgroup of $\{H, U_3, U_4\}$ is of order p ; let it be generated by s_5 . The commutators arising from transformation of s_1 and s_2 by U_1 may be assumed to be independent of s_5 , otherwise $\{U_1, U_3, U_4\}$ would contain operators of type I not in $\{U_3, U_4\}$. These commutators may then be taken to be s_3 and s_4 respectively. Then U_2 may be chosen so that the commutator subgroup of $\{H, U_2\}$ is in $\{s_3, s_4\}$. If $\{U_1, U_2\}$ contains no operator of type I, as must be the case, generators of G satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_4, & U_3^{-1}s_1U_3 &= s_1s_5, \\ U_1^{-1}s_2U_1 &= s_2s_4, & U_2^{-1}s_2U_2 &= s_2s_3, & U_4^{-1}s_2U_4 &= s_2s_5. \end{aligned}$$

In case (b) we may assume that generators of G satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_5, & U_3^{-1}s_1U_3 &= s_1s_4, \\ U_1^{-1}s_2U_1 &= s_2s_j, & U_2^{-1}s_2U_2 &= s_2s_3, & U_4^{-1}s_2U_4 &= s_2s_5. \end{aligned}$$

For the commutator subgroup of $\{H, U_1, U_2\}$ is of order at most p^3 , and the commutator subgroups of order p^2 arising from transformation of s_1 and s_2 respectively must have an operator in common which does not belong to $\{s_4, s_5\}$. Each of the groups H_1 and H_2 must be of order p^3 . Therefore the commutator of U_2 and s_1 can be taken to be s_5 . The operator s_j is in $\{s_3, s_4, s_5\}$ and can be assumed to be $s_3^\alpha s_4^\beta$. If α is not zero, then $\{U_1, U_3\}$ contains an operator of type I which is not a power of U_3 . We may assume $\alpha = 0$ and $\beta = 1$, in which case $U_1U_2U_3U_4$ is of type I. Therefore there is no group satisfying these conditions.

When $l = 2$, both H_1 and H_2 are of order p^2 . It is therefore possible to find generators of G which satisfy the relations:

$$\begin{aligned} U_1^{-1}s_1U_1 &= s_1s_3, & U_2^{-1}s_1U_2 &= s_1s_4, \\ U_3^{-1}s_2U_3 &= s_2s_3, & U_4^{-1}s_2U_4 &= s_2s_4. \end{aligned}$$

Hence there is but one group for $l = 2$.

We proceed to list the groups with enough information to determine the group in each case. The first column contains the value of l , which is all the information necessary for the first group and the last. The second column gives the number of subgroups of type I in U . The third column contains whatever further information may be necessary. This additional information in most cases takes the form of a statement of the existence or the non-existence in G of a "subgroup of order p^{n+a} with commutator subgroup of order p^β and no operator of type I"; for such a subgroup we shall use the symbol $G_{a,\beta}$. Since it cannot lead to any confusion we shall use the symbol $G_{1,1}$ to stand for a subgroup of type I.

l	$G_{1,1}$'s	other facts	l	$G_{1,1}$'s	other facts
1. 8	—	—	23.	0	a $G_{3,3}$, no $G_{2,2}$
2. 7	1	—	24.	0	no $G_{2,2}$, or $G_{3,3}$
3.	0	—	25. 4	$2 + p + p^2$	—
4. 6	$1 + p$	—	26.	$2 + 2p$	—
5.	2	—	27.	$3 + p$	—
6.	1	a $G_{3,5}$	28.	4	—
7.	1	no $G_{3,5}$	There are $(p+1)/6$ or $(p+5)/6$ such groups.		
8.	0	a $G_{2,2}$	29. 4	$1 + p$	a $G_{2,2}$
9.	0	a $G_{3,4}$ and no $G_{2,2}$	30.	2	a $G_{2,2}$
10.	0	no $G_{2,2}$, no $G_{3,4}$	There are $(p+1)/2$ such groups.		
11. 5	$1 + p + p^2$	—	31. 4	1	—
12.	$2 + p$	—	There are $(p+1)/3$ or $(p+5)/3$ such groups.		
13.	3	—	32. 4	0	—
14.	$1 + p$	a $G_{2,3}$	There are $p+2$ such groups. (Cf. above.)		
15.	$1 + p$	no $G_{2,3}$	33. 3	$1 + 2p + p^2$	—
16.	$1 + p$	¹¹	34.	$1 + 3p$	—
17.	2	a $G_{2,3}$	35.	3	—
18.	2	no $G_{2,3}$	36.	$1 + p$	—
19.	1	a $G_{2,2}$	37. 2	—	—
20.	1	a $G_{3,4}$, no $G_{2,2}$			
21.	1	no $G_{2,2}$, no $G_{3,4}$			
22. 5	0	a $G_{2,2}$			

There are in all $2p + 36$ or $2p + 38$ distinct groups according as p is of the form $6k - 1$ or $6k + 1$.

¹¹ Contains a subgroup of order p^{n+2} with commutator subgroup of order p ; the two preceding groups have no such subgroup.

2. *The bilinear forms.* As was explained for the case $l = 4$, the commutator structure of G can be described by means of two matrices M and N . In the general case these are matrices of four rows and l columns. That there are two matrices depends on the fact that we are discussing groups G whose centrals are of order p^{n-2} , and therefore generators of H can be chosen so that all but two are in the central of G . The argument in the paper cited above for $l = 4$ holds for any l . A change in the generators of the commutator subgroup of G , when the U 's and s_1 and s_2 are not changed, replaces the columns of M by linear combinations of its columns, and replaces the columns of N by the same linear combinations of its columns. The effect is to replace M and N respectively by MB and NB where B is a non-singular square matrix of l rows. A change in the generators of U has the effect of replacing M and N respectively by AM and AN where A is a non-singular four-rowed square matrix. A change in the generators of $\{s_1, s_2\}$ has the effect of replacing the matrix $\lambda_1 M + \lambda_2 N$ by the matrix $(a\lambda'_1 + b\lambda'_2)M + (c\lambda'_1 + d\lambda'_2)N$ where $s'_1 = s_1^a s_2^b$, $s'_2 = s_1^c s_2^d$, and $(ad - bc) \not\equiv 0$.

The matrices M and N may be interpreted as the matrices of two bilinear forms in four variables x_1, x_2, x_3, x_4 and l variables y_1, y_2, \dots, y_l , in which case $\lambda_1 M + \lambda_2 N$ becomes the matrix of a member of the pencil of bilinear forms determined by the two whose matrices are M and N . The changes on generators of the groups then correspond to linear homogeneous non-singular transformations on the x 's, the y 's, and the λ 's. The problem of classification of the groups is then identically the problem of classification of pencils of bilinear forms for $l = 2, \dots, 8$ under these transformations, for every change of generators of G gives a transformation of the pencil and every transformation of the pencil, with coefficients in the modular field, gives a transformation on generators of the group. In the case of $l = 4$ we were able to classify the groups most easily by means of the theory of invariant factors of $\lambda_1 M + \lambda_2 N$. when M and N are not square a corresponding theory has not been developed. There are some obvious difficulties in the way of extending the theory for square matrices.

Having classified the group for these various values of l we are able to write down immediately a set of normal forms for pencils of bilinear forms in four x 's and l y 's. We give these normal forms here numbered in the same way as the groups at the end of § 1.

1. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_5 + x_2y_6 + x_3y_7 + x_4y_8).$
2. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_5 + x_2y_6 + x_3y_7).$
3. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_5 + x_2y_6 + x_3y_7 + x_4y_1).$
4. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_5 + x_2y_6).$

5. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2(x_1y_4 + x_2y_5 + x_4y_6).$
6. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2(x_1y_4 + x_2y_5 + x_3y_2 + x_4y_6).$
7. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2(x_1y_4 + x_2y_5 + x_3y_6 + x_4y_3).$
8. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_2 + rx_2y_1 + x_3y_5 + x_4y_6),$
 r not a square.
9. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_5 + x_2y_1 + x_3y_2 + x_4y_6).$
10. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_5 + x_2y_6 + x_3y_1 + x_4y_2).$
11. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2x_1y_5.$
12. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2(x_1y_4 + x_4y_5).$
13. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2(x_1y_4 + x_3y_3 + x_4y_5).$
14. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_5 + x_2y_1).$
15. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_5 + x_2y_3).$
16. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2(x_1y_4 + x_2y_5 + x_4y_5).$
17. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2(x_1y_4 + x_2y_1 + x_4y_6).$
18. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2(x_1y_4 + x_2y_3 + x_4y_5).$
19. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_2 + rx_2y_1 + x_3y_5).$
20. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_5 + x_2y_1 + x_3y_2).$
21. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_5 + x_2y_1 + x_3y_4).$
22. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_2 + rx_2y_1 + x_3y_5 + x_4y_3).$
23. $\lambda_1(x_1y_1 + x_2y_2 + x_3[\alpha y_1 + \beta y_3] + x_4y_4) + \lambda_2(x_1y_3 + x_2y_1 + x_3y_2 + x_4y_5).$
24. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_5 + x_2y_1 + x_3y_2 + x_4y_3).$
25. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2x_4y_4.$
26. $\lambda_1(x_1y_1 + x_2y_2) + \lambda_2(x_3y_3 + x_4y_4).$
27. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2(x_3y_3 + x_4y_4).$
28. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2(x_2y_2 + \rho x_3y_3 + x_4y_4).$
29. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_2 + rx_2y_1).$
30. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2(x_1y_2 + rx_2y_1 + x_4y_4).$
31. $\lambda_1(x_1y_1 + x_2y_2 + x_3[\alpha y_1 + \beta y_3] + x_4y_4) + \lambda_2(x_1y_3 + x_2y_1 + x_3y_2).$
32. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4) + \lambda_2(x_1y_2 + rx_2y_1 + x_3y_4 + rx_4y_3).$
 $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3 + x_4[\alpha y_4 + \beta y_1 + \gamma y_2 + \delta y_3])$
 $+ \lambda_2(x_1y_4 + x_2y_1 + x_3y_2 + x_4y_3),$
- where $\lambda^4 + \delta\lambda^3 - \gamma\lambda^2 + \beta\lambda - \alpha$ has no linear factor.
33. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2x_4y_3.$
34. $\lambda_1(x_1y_1 + x_2y_2) + \lambda_2(x_3y_2 + x_4y_3).$
35. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2(x_1[y_1 + y_3] + x_3y_3 + x_4y_1).$
36. $\lambda_1(x_1y_1 + x_2y_2 + x_3y_3) + \lambda_2(x_1y_2 + rx_2y_1 + x_4y_3).$
37. $\lambda_1(x_1y_1 + x_2y_2) + \lambda_2(x_3y_1 + x_4y_2).$

It is interesting to interpret in terms of the bilinear forms the considera-

tions that were used in classifying the groups. The separation of the groups into classes according to orders of their commutator subgroups corresponds to the separation of the bilinear forms into classes according to the number of the variables y . One of the first things which attracts attention with regard to the groups is that if U is of order p^4 the order of the commutator subgroup cannot be greater than p^3 . This says that any pencil of bilinear forms in four x 's and l y 's is expressible as a pencil of forms in four x 's and l' y 's where $l' \leq 8$. This is of course obvious from a consideration of the matrices of the forms in terms of which the pencil is expressed.

A pair of numbers λ_1, λ_2 determines a bilinear form of the given pencil; the pair also determines an operator $s_1^{\lambda_1} s_2^{\lambda_2}$ of the group $\{s_1, s_2\}$. When these particular values λ_1, λ_2 are substituted in the expression $\lambda_1 M + \lambda_2 N$ the resulting matrix has four rows each determining an operator of the commutator subgroup arising from transformation of $s_1^{\lambda_1} s_2^{\lambda_2}$ by U . If U contains an operator permutable with $s_1^{\lambda_1} s_2^{\lambda_2}$ the rank of this matrix is less than four. The corresponding bilinear form determined by λ_1, λ_2 will therefore have a rank less than four. Consequently the separation of the groups with commutator subgroups of a given order into classes according to the number of operators of type I in U corresponds to the separation of the pencils of bilinear forms with four x 's and l y 's into classes according to the number of forms in the pencils which have ranks less than four. We may call such a form, of rank less than four, *singular*. Then the classification of pencils has been made according to the number of singular forms in a given pencil. Theorem (1.1) states that *any two pencils of bilinear forms in m x 's and $2m$ y 's are conjugate and that such a pencil contains no singular form*. Theorem (1.2) states that *there are two distinct pencils of bilinear forms in m x 's and $2m - 1$ y 's; one contains no singular form, and the other contains one*.¹²

When $l = 6$ the classification according to the number of singular forms in the pencil is not enough. There are two pencils which have each one singular form. In both 6 and 7 above the singular form appears for $\lambda_1, \lambda_2 = 1, 0$. We may distinguish between them in the following way: Consider the numbers (x_1, x_2, x_3, x_4) to be the coördinates of a point in a finite three-space. In terms of the coördinates of the plane $x_4 = 0$ no. 6 determines a pencil of forms in three x 's and five y 's and no. 7 determines a pencil of forms in three x 's and six y 's. The first pencil contains no singular form. It is possible to select planes in (x_1, x_2, x_3, x_4) in terms of whose coördinates

¹² Where we understand the forms $F(x_1, \dots, y_l)$ and $k \cdot F(x_1, \dots, y_l)$ to be the same; otherwise the number is $p - 1$.

no. 7 will give pencils of forms in three x 's and five y 's, but every such pencil will contain singular forms.

There are three distinct pencils of forms in four x 's and six y 's none of which contains a singular form. None of the pencils of forms in the variables of a subspace of (x_1, x_2, x_3, x_4) can be singular. No. 8 determines a pencil of forms in two x 's and two y 's on the line $x_3 = x_4 = 0$. There is no line on which nos. 9 or 10 determines such a pencil. No. 9 determines a pencil of forms on three x 's and four y 's on the plane $x_4 = 0$, and there is no plane on which no. 10 determines such a pencil.

The interpretation in terms of bilinear forms is particularly enlightening in the case of no. 28. There are $(p+1)/6$ or $(p+5)/6$ such pencils depending on the value of p . Each pencil contains four singular forms, each singular form is given by a pair λ_1, λ_2 . Each form of the pencil determines a point on the line (λ_1, λ_2) . The cross-ratio of these four points remains invariant under projective transformation of the ordered set of four points. A reordering of the four points gives one of six values of the cross-ratio. Two forms determined by ρ and ρ' cannot be conjugate unless ρ' is one of the values $\rho, 1-\rho, 1/\rho, 1/(1-\rho), (\rho-1)/\rho$, or $\rho/(\rho-1)$.

The differences among the groups that come under 30, 31, and 32 can all be interpreted in terms of pencils induced in subspaces of (x_1, x_2, x_3, x_4) . It is suggested that perhaps a more thoroughgoing geometric interpretation of the whole situation would be worth while.

3. *The proof of distinctness.* We come now to the question of the possible isomorphism of two groups belonging to different classes. Our proofs of uniqueness of the various normal forms of generating relations have always been proofs of uniqueness under automorphisms of G in which H corresponds to itself. In none of these groups has H been the only abelian group of order p^n in G , for the group $\{U_1, U_2, s_3, s_4, \dots, s_n\}$ is such a group. Moreover, the group U is not in general a characteristic subgroup of G , for any operator U_i may be replaced by sU_i where s is any operator of H . If s is in the central of G this has no effect on generating relations, but if s is not in the central the U 's will in general cease to be permutable. We are confining our attention to groups G in which the U 's are permutable. This is justified by the fact that any classification of metabelian groups must take these groups into account; it must depend on the possibilities of "commutator structure" arising from transformation of H by U . The simplicity of the statement in terms of pencils of bilinear forms gives added assurance that the limitations imposed on the investigations are natural to the problem and not arbitrary.

The question may be considered from the point of view of the uniqueness of the defining relations. The particular defining relations that we have chosen in each case are not the only ones that could be chosen. The situation is rather that if certain properties of the generators are required, then the defining relations can be reduced to the normal form we have given for the particular group. If two groups G and G' belonging to different classes were simply isomorphic, then it would be possible to select generators of G' which satisfied defining relations of G . These generators would in every case satisfy the following conditions:

- (1) n of the generators would generate an abelian subgroup H' , and and $n - 2$ of them would generate the central of G' ;
- (2) the remaining four would generate an abelian group of type 1, 1, 1, 1;
- (3) no operator of the second group, except identity, would be permutable with every operator of the first;
- (4) the operators of the second group would correspond to operators of types I and II in the I-group of the first.

We have already considered the possibility of such isomorphisms between G and G' in which H corresponds to itself and have seen that none exists when the normal forms of the generating relations are different. The group H' cannot then be $\{s_1, s_2, \dots, s_n\}$; it must, however, include the central $\{s_3, s_4, \dots, s_n\}$. We assume then that G and G' have different normal forms for their generating relations and that there exists a simple isomorphism between them. Then there exists in G a group H' which corresponds to H in G' . This group H' must be $\{s'_1, s'_2, s_3, s_4, \dots, s_n\}$, where

$$s'_i = s_1^{a_i} s_2^{\beta_i} U_1^{k_i} U_2^{l_i} U_3^{m_i} U_4^{n_i}, \quad (i = 1, 2).$$

The numbers k_1, \dots, n_2 cannot all be zero for then H' would be H . We may suppose that all but k_1 and l_2 or all but k_1 and k_2 are zero, for this requires only a proper choice of generators of U . We may suppose further that the two which are not zero are both equal to 1, so that $s'_1 = s_1^{a_1} s_2^{\beta_1} U_1$ and $s'_2 = s_1^{a_2} s_2^{\beta_2} U_i$, ($i = 1$ or 2). If $\alpha_1 \beta_2 - \alpha_2 \beta_1 \neq 0$ we may select generators of $\{s_1, s_2\}$ so that $s'_1 = s_1 U_1$ and $s'_2 = s_2 U_i$. If then $i = 1$, we may replace s'_2 by $s_2'' = s_2$. If $i = 2$, the group $U' = \{U'_1, U'_2, U'_3, U'_4\}$ where

$$U'_i = s_1^{a_i} s_2^{\beta_i} U_1^{k_i} U_2^{l_i} U_3^{m_i} U_4^{n_i}$$

must contain U_1 or s_1 and U_2 or s_2 . U' cannot contain both s_1 and s_2 , for then without changing generating relations we could replace H' by

$H'' = \{U_1, U_2, s_3, s_4, \dots, s_n\}$ which is not maximal abelian invariant. We may therefore suppose that $s'_1 = s_1 U_1$ and $s'_2 = s_2$. If on the other hand $\alpha_1 \beta_2 - \alpha_2 \beta_1 = 0$, we may suppose that $s'_1 = s_1 U_1$ and $s'_2 = U_i$. If $U_i = U_1$, s_1 is in H' and we may take s'_1 to be s_1 . If U_i is not U_1 , then either U_1 or s_1 is in U' . We have seen that there are not two independent U 's in H' and therefore whatever the value of $\alpha_1 \beta_2 - \alpha_2 \beta_1$ we may suppose that

$$s'_1 = s_1 U_1 \quad \text{and} \quad s'_2 = s_2.$$

From this it follows that U_1 must be permutable with s_2 and hence must be of type I. In the expression for one of the U'_i 's as given above in terms of $s_1, s_2, U_1, \dots, U_4$ the exponent μ_i must be different from zero. Making use of it, s'_1 may be replaced by $s_1'' = U_1$ without affecting the generating relations. The group U' must then contain s_1 or else by the same sort of transformation H' can be changed to H . But if s_1 is in U' it must be permutable with U_2, U_3 , and U_4 . This identifies the group as no. 25. Since this is the only one of the groups with $2 + p + p^2$ subgroups of type I, the original sets of relations of generators of G and G' were transformable into each other.

4. *The general case.* The theorems (1.1), (1.2), and (1.4) go beyond the case where the order of U is p^4 . It is clear that the methods used will suffice to classify the groups G of order p^{n+m} where U of order p^m and abelian of type 1, 1, \dots contains only operators of types I and II. It is clear also that this problem is the same as the problem of classification of families of forms. The groups of order p^{n+m} may be separated first into classes according to the orders of their commutator subgroups. Then each of these classes may be subdivided according to the number of independent operators of type I in U . When these U 's of type I are segregated, the remaining operators of a set of independent generators of U determine a group U' whose operators are all of type II. U' determines two groups H'_1 and H'_2 which are commutator subgroups arising from transformation of s_1 and s_2 respectively by U' . Unless the commutator subgroup of $\{H, U'\}$ is the product of the two distinct groups H'_1 and H'_2 , H'_1 and H'_2 have a cross-cut different from identity. This cross-cut determines operators U_1, U_2, \dots such that their commutators with s_1 generate the cross-cut and it determines operators V_1, V_2, \dots such that their commutators with s_2 generate the cross-cut. The order of the group $U'' = \{U_1, U_2, \dots, V_1, V_2, \dots\}$ enables us to determine a normal form for the relations defining $\{H, U''\}$ and then a normal form for relations defining G . The kinds of differences that may present themselves are apparent. For example, the operators $U_1, U_2, \dots, V_1, V_2, \dots$ may be independent or they

may be dependent in various ways. It may be possible to select α U 's and α V 's such that the commutator subgroup determined by them and H is of order p^α , in which case G contains a subgroup $G_{\alpha,\alpha}$. If that is the case it is necessary to determine the type of $G_{\alpha,\alpha}$ which appears. This goes back to the question of the invariant factors of the matrix $\lambda_1 M + \lambda_2 N$ where M and N are α -rowed square matrices. Though the method is clear and obviously sufficient it would be desirable to continue the study on account of the interesting facts that are bound to appear in the classification of polynomials of degree even as small as five.

It is likewise obvious that the methods used here are sufficient for the classification of groups where U contains operators of type more "advanced" than I and II. If U contains an operator of type III, one which determines the partition $n = 2 + 2 + 2 + 1 + \dots + 1$, and no operator except those of types I, II, and III, then the central of G would be of order p^{n-3} . Only three of the generators of H , s_1 , s_2 , and s_3 , need be outside the central of G . They would determine three matrices M_1 , M_2 , M_3 and three bilinear forms. The classes of groups would correspond in a 1 — 1 manner with the classes of three-parameter (homogeneous) families of bilinear forms $\lambda_1 M_1 + \lambda_2 M_2 + \lambda_3 M_3$. This classification would probably depend on the types of pencil as well as the types of form contained in the family. The method of procedure is clear in its general aspects; the details of the possible difficulties are not so clear. On that account the classification should be carried on in detail somewhat further in this direction.

The methods and results point the way also to the treatment of groups where U contains operators of type \bar{K} and none of type greater than \bar{K} . We shall content ourselves with the statement of the following theorem which has been clear for some time, although the theorem does not take full account of the method of attack we have used.

The problem of the classification of metabelian groups $\{H, U\}$ of order p^{n+m} which contain H as a maximal invariant abelian group and in which U is in the group of isomorphisms of H is identical with the problem of classification of k -parameter (homogeneous) families of bilinear forms in m variables x and an undetermined number of variables y under "rational" projective transformation on the x 's, the y 's, and on the parameters. The number k takes on all values not greater than $n/2$.

A METRICALLY TRANSITIVE GROUP DEFINED BY THE MODULAR GROUP.

By GUSTAV A. HEDLUND.¹

1. Introduction. It is known ² that the modular group, Γ , is metrically transitive with respect to the real axis. This means that if E is a measurable point set of the real axis which is invariant under the transformations of the modular group, Γ , either E or its complement with respect to the real axis is a zero set.

Let Γ_2 be the group of real transformations of the (ξ, η) plane, π , into itself, given by

$$T: \quad \xi = \frac{a\xi' + b}{c\xi' + d}, \quad \eta = \frac{a\eta' + b}{c\eta' + d},$$

where a, b, c and d are integers such that $ad - bc = 1$. The object of this paper is to prove that the group Γ_2 is metrically transitive with respect to the plane π .

This is the essential result needed in proving the metrical transitivity of the dynamical system obtained by considering the non-euclidean billiard problem ³ defined by the modular group. The similar result where the modular group is replaced by a certain Fuchsian group with closed fundamental region has been published by the author.⁴ The technic in the two cases is similar, but in the present case the method is not buried under the details involved in the other case.

The result obtained here implies the metrical transitivity of the group Γ with respect to the real axis, but the proof is, of course, very indirect. Conversely, as an example shows, metrical transitivity of a group G with respect to the real axis does not necessarily imply metrical transitivity of the group G_2 , which is obtained by applying the transformations of G simultaneously to two variables, with respect to the plane of the two variables.

¹ This paper was completed during the tenure of a National Research Fellowship.

² M. H. Martin, "Metrically Transitive Point Transformations," *Bulletin of the American Mathematical Society*, vol. 40 (1934), pp. 606-612.

³ E. Artin, "Ein mechanisches System mit quasiergodischen Bahnen," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, vol. 3 (1924), pp. 170-175.

⁴ G. A. Hedlund, "On the metrical transitivity of the geodesics on closed surfaces of constant negative curvature," *Annals of Mathematics*, vol. 35 (1934), pp. 787-808.

2. Quasi-transitive points. Under the transformations of the group Γ_2 a point P of π is transformed into the set of points congruent to P . If this set is everywhere dense in π the point P will be called a *quasi-transitive* point. From the work of Artin ⁵ it is known that not only are there quasi-transitive points, but a point is quasi-transitive if either its abscissa or ordinate belongs to a certain linear set, the complement of which with respect to the real axis is a linear zero set. It immediately follows that the set of non-quasi-transitive points of π constitutes a zero set.

With the aid of this result the problem of proving that any invariant measurable set in π is either a zero set or the complement of a zero set is considerably simplified. For since the non-quasi-transitive points form an invariant zero set, these points can be omitted from a measurable invariant set without affecting either the measure or the invariance. The use of this fact is illustrated in the following lemma.

LEMMA 2.1. *If E is an invariant measurable set of π and D is an open set of π , E is a zero set if $E \cdot D$ is a zero set.*

For assuming all points of E quasi-transitive, the set E can be obtained from the set $E \cdot D$ by applying the transformations of the group Γ_2 . But if $E \cdot D$ is a zero set, the set obtained by applying the denumerable set of transformations of Γ_2 is a zero set.

In particular the set D will be chosen as the set $1 < \xi < 2$, $-1 < \eta < 0$. If E is the given invariant measurable set and it is shown that either $E \cdot D$ or $D - E \cdot D$ is a zero set, the desired theorem will have been proved.

3. A net in D . Let $P(\xi, \eta)$ be a point of D and let the developments of ξ and $-\eta$ in continued fractions with positive integral partial quotients be given by

$$\xi = [1, a_1, a_2, \dots], \quad -\eta = [0, a_{-1}, a_{-2}, \dots],$$

where $[b_0, b_1, b_2, \dots]$ is given by ⁶

$$[b_0, b_1, \dots] = b_0 + \frac{1}{b_1 + \frac{1}{b_2 + \text{etc.}}}.$$

and the continued fraction may or may not be terminating. If it is terminating, the representation will not be unique, but this does not affect the following arguments.

⁵ E. Artin, *loc. cit.*, p. 174.

⁶ Perron, *Die Lehre von den Kettenbrüchen*, Teubner, 1913, p. 27.

Given the two sets of positive integers, a_1, a_2, \dots, a_μ , and $a_{-1}, a_{-2}, \dots, a_{-\nu}$, let R be those points (ξ, η) of D such that $\xi = [1, a_1, a_2, \dots, a_\mu, \dots]$ and $-\eta = [0, a_{-1}, a_{-2}, \dots, a_{-\nu}, \dots]$, where again these may be terminating continued fractions but they must contain enough partial quotients to assure the presence of the given sequences. The set R includes all the points of a rectangle in D . The interior, Δ , of R , will be denoted by

$$\{1, a_1, \dots, a_\mu; 0, a_{-1}, \dots, a_{-\nu}\}.$$

Thus, in particular, D is given by $\{1; 0\}$.

Let $L(a_1, a_2, \dots, a_\mu)$ be the length of a side of Δ parallel to the x -axis and $L(a_{-1}, a_{-2}, \dots, a_{-\nu})$ the length of a vertical side. The following formula is readily obtained:

$$(3.1) \quad L(a_1, \dots, a_\mu) = \prod_{i=1}^{\mu} (P_i P'_i)^{-1},$$

where $P_i = [a_i, \dots, a_\mu]$ and $P'_i = [a_i, \dots, a_{\mu-1}, a_\mu + 1]$, ($i = 1, 2, \dots, \mu$). A similar formula holds for $L(a_{-1}, \dots, a_{-\nu})$.

4. A lemma on continued fractions.

LEMMA 4.1.

$$F(x_0, x_1, \dots, x_n) = \frac{[x_0, x_2, \dots, x_n]}{[x_0, x_1, \dots, x_n + \alpha]}, \quad 0 \leq \alpha, \quad 1 \leq x_i < \infty.$$

If n is even (odd), F is a non-decreasing (non-increasing) function of each of the variables x_i , ($i = 0, 1, \dots, n$).

Let $P_i = [x_i, \dots, x_n]$ and $Q_i = [x_i, \dots, x_{n-1}, x_n + \alpha]$. Then obviously, $\partial P_0 / \partial x_0 = \partial Q_0 / \partial x_0 = 1$, and the following formulas can be obtained by evaluating the limit of the difference quotient:

$$(4.2) \quad \frac{\partial P_0}{\partial x_i} = (-1)^i \prod_{k=1}^i P_k^{-2}; \quad \frac{\partial Q_0}{\partial x_0} = (-1)^i \prod_{k=1}^i Q_k^{-2}; \quad (i = 1, 2, \dots, n).$$

From these formulas follow:

$$(4.3) \quad \frac{\partial F}{\partial x_0} = \frac{Q_0 - P_0}{Q_0^2}; \quad \frac{\partial F}{\partial x_i} = (-1)^i Q_0^{-1} \left\{ 1 - \frac{P_0}{Q_0} \prod_{k=1}^i \frac{P_k^2}{Q_k^2} \right\} \prod_{k=1}^i P_k^2, \\ (i = 1, 2, \dots, n).$$

Case I, n even. Then $P_i \leq Q_i$, i even, and $P_i \geq Q_i$, i odd. In this case it follows at once from (4.3) that $\partial F / \partial x_0 \geq 0$. From (4.3) we have

$$\frac{\partial F}{\partial x_1} = -Q_0^{-1} P_1^{-2} \left\{ 1 - \frac{P_0}{Q_0} \frac{P_1^2}{Q_1^2} \right\}.$$

But

$$\frac{P_0 P_1}{Q_0 Q_1} = \frac{x_0 P_1 + 1}{x_0 Q_1 + 1} \geq 1,$$

since $P_1 \geq Q_1$, and hence $\partial F / \partial x_1 \geq 0$. To complete the proof by induction, we assume the theorem true for $(i = 1, 2, \dots, j < n)$. Then if j is even we wish to show that $P_0 Q_0^{-1} \prod_{k=1}^{j+1} P_k^2 Q_k^{-2} \geq 1$. From the assumption that the result holds for $i < j + 1$, it follows that $P_0 Q_0^{-1} \prod_{k=1}^{j-1} P_k^2 Q_k^{-2} \geq 1$. But the inequality $P_{j+1} \geq Q_{j+1}$ implies that $P_j P_{j+1} / Q_j Q_{j+1} = (x_j P_{j+1} + 1) / (x_j Q_{j+1} + 1) \geq 1$, and the desired result follows. The case where j is odd is treated similarly.

Case II, n odd. Then $P_i \geq Q_i$, i even, $P_i \leq Q_i$, i odd, and from the reversal of the inequalities, the proof in this case follows readily from that given in Case I.

5. The fundamental lemma.

LEMMA 5.1. *There exist positive constants k_1 and k_2 such that*

$$\begin{aligned} k_1 \frac{L(a'_1, \dots, a'_m, a_{m+1}, \dots, a_n)}{L(a'_1, \dots, a'_m)} &< \frac{L(a_1, \dots, a_m, a_{m+1}, \dots, a_n)}{L(a_1, \dots, a_m)} \\ &< k_2 \frac{L(a'_1, \dots, a'_m, a_{m+1}, \dots, a_n)}{L(a'_1, \dots, a'_m)}, \quad n \geq m > 0, \end{aligned}$$

independent of the positive integers $a'_1, \dots, a'_m, a_1, \dots, a_m, a_{m+1}, \dots, a_n$.

Proof. Let

$$\begin{aligned} P_i &= [a_i, \dots, a_n]; & P'_i &= [a_i, \dots, a_n + 1], \quad (i = 1, \dots, n); \\ Q_i &= [a'_i, \dots, a'_m, a_{m+1}, \dots, a_n]; & Q'_i &= [a'_i, \dots, a'_m, a_{m+1}, \dots, a_n + 1], \\ & & & (i = 1, \dots, m); \\ R_i &= [a_i, \dots, a_m]; & R'_i &= [a_i, \dots, a_m + 1], \quad (i = 1, \dots, m); \\ S_i &= [a'_i, \dots, a'_m]; & S'_i &= [a'_i, \dots, a'_m + 1], \quad (i = 1, \dots, m). \end{aligned}$$

Then

$$\begin{aligned} L(a_1, \dots, a_m, a_{m+1}, \dots, a_n) &= \prod_{i=1}^n (P_i P'_i)^{-1}; \\ L(a'_1, \dots, a'_m, a_{m+1}, \dots, a_n) &= \prod_{i=1}^m (Q_i Q'_i)^{-1} \prod_{i=m+1}^n (P_i P'_i)^{-1}; \\ L(a_1, \dots, a_m) &= \prod_{i=1}^m (R_i R'_i)^{-1}; \\ L(a'_1, \dots, a'_m) &= \prod_{i=1}^m (S_i S'_i)^{-1}; \end{aligned}$$

and

$$\frac{L(a_1, \dots, a_m) L(a'_1, \dots, a'_m)}{L(a_1, \dots, a_m) L(a'_1, \dots, a'_m, a_{m+1}, \dots, a_n)} = \prod_{i=1}^m \frac{R_i R'_i Q_i Q'_i}{P_i P'_i S_i S'_i} \\ = G(a_1, \dots, a_m, a'_1, \dots, a'_m, a_{m+1}, \dots, a_n).$$

Let us consider

$$\frac{R_i}{P_i} = \frac{[a_i, \dots, a_m]}{[a_i, \dots, a_m, a_{m+1}, \dots, a_n]}.$$

From Lemma 4.1, if $m - i$ is even,

$$\frac{R_i(1)}{R_{i-1}(1)} \leq \frac{R_i}{P_i} \leq 1,$$

where $R_i(1)$ is the number obtained by replacing each argument in R_i by 1. Similarly, if $m - i$ is odd

$$1 \leq \frac{R_i}{P_i} \leq \frac{R_i(1)}{R_{i-1}(1)}.$$

Hence

$$\left(\frac{R_m(1)}{R_{m-1}(1)} \cdot \frac{R_{m-2}(1)}{R_{m-3}(1)} \cdot \dots \right) \prod_{i=1}^m \frac{R'_i Q_i Q'_i}{P'_i S_i S'_i} \\ \leq G \leq \left(\frac{R_{m-1}(1)}{R_{m-2}(1)} \frac{R_{m-3}(1)}{R_{m-4}(1)} \cdot \dots \right) \prod_{i=1}^m \frac{R'_i Q_i Q'_i}{P'_i S_i S'_i}$$

where the products in the parentheses are continued as long as the subscripts remain positive. From this follows

$$\left(\frac{[1]}{[1, 1]} \frac{[1, 1, 1]}{[1, 1, 1, 1]} \cdot \dots \right) \prod_{i=1}^m \frac{R'_i Q_i Q'_i}{P'_i S_i S'_i} \\ \leq G \leq \left(\frac{[1, 1]}{[1, 1, 1]} \frac{[1, 1, 1, 1]}{[1, 1, 1, 1, 1]} \cdot \dots \right) \prod_{i=1}^m \frac{R'_i Q_i Q'_i}{P'_i S_i S'_i},$$

provided the infinite products in the parentheses converge. But the successive convergents of $(1 + 5^{1/2})/2$ are given by $c_1 = [1]$, $c_2 = [1, 1]$, \dots , and it is readily shown that the infinite products $\prod_{i=1}^{\infty} c_{2i-1}/c_{2i}$ and $\prod_{i=1}^{\infty} c_{2i}/c_{2i+1}$ converge to positive constants e_1 and e_2 , respectively.

Using the same technic it can be shown that the desired inequality

$$e_1^2/e_2^2 \leq G \leq e_2^2/e_1^2$$

obtains. Setting $k_1 = e_1^2/e_2^2$ and $k_2 = e_2^2/e_1^2 = k_1^{-1}$, the desired lemma holds.

6. Some inequalities. Let $\Delta = \{1, a_1, \dots, a_\mu; 0, a_{-1}, a_{-2}, \dots, a_{-v}\}$ be a chosen rectangle of the net in D . Let σ_λ denote the set of sub-rectangles

of Δ given by $\{1, a_1, \dots, a_\mu, s_1, s_2, \dots, s_\lambda, 1; 0, a_{-1}, a_{-2}, \dots, a_{-v}\}$, where $s_1, s_2, \dots, s_\lambda$ are arbitrary positive integers, but λ is so chosen that $\lambda + \mu$ is odd. The area of the set σ_λ is denoted by $\mu(\sigma_\lambda)$.

LEMMA 6.1. *There exists a positive constant k_3 such that*

$$\mu(\sigma_\lambda) > k_3 \mu(\Delta).$$

This amounts simply to proving that there exists a positive constant k_3 such that

$$L(a_1, a_2, \dots, a_m, 1) > k_3 L(a_1, a_2, \dots, a_m),$$

where k_3 is independent of a_1, a_2, \dots, a_m .

A brief computation with the aid of (3.1) yields

$$\begin{aligned} & \frac{L(a_1, a_2, \dots, a_m, 1)}{L(a_1, a_2, \dots, a_m)} \\ &= \frac{[a_1, a_2, \dots, a_m]}{[a_1, a_2, \dots, a_m + 1/2]} \cdot \frac{[a_2, \dots, a_m]}{[a_2, \dots, a_m + 1/2]} \cdots \frac{a_m}{a_m + 1/2} \cdot \frac{1}{2}. \end{aligned}$$

With the aid of Lemma 4.1 and using the notation of the preceding paragraph,

$$\frac{L(a_1, a_2, \dots, a_m, 1)}{L(a_1, a_2, \dots, a_m)} \geq \frac{1}{c_2} \cdot \frac{1}{c_3} \cdot \frac{c_3}{c_5} \cdot \frac{c_5}{c_7} \cdots = \frac{1}{1 + 5^{1/2}}$$

and the above lemma holds with $k_3 = (1 + 5^{1/2})$.

Using the fact that $\lambda + \mu$ is odd, the rectangle

$$\bar{\Delta} = \{1, a_1, a_2, \dots, a_\mu, s_1, \dots, s_\lambda, 1; 0, a_{-1}, a_{-2}, \dots, a_{-v}\}$$

is defined by the inequalities:

$$\begin{aligned} \xi_1 = [1, a_1, \dots, a_\mu, s_1, \dots, s_\lambda, 1] &< \xi < [1, a_1, \dots, a_\mu, s_1, \dots, s_\lambda, 2] = \xi_2, \\ -\eta_1 = [0, a_{-1}, \dots, a_{-v}] &< -\eta < [0, a_{-1}, \dots, a_{-v} + 1] = -\eta_2, \end{aligned}$$

provided v is odd. The second inequality is reversed if v is even, but it will be sufficient to discuss the case v odd.

The transformation $\xi = [1, a_1, \dots, a_\mu, s_1, \dots, s_\lambda, \xi']$ is given by

$$(6.1) \quad \xi = \frac{a\xi' + b}{c\xi' + d}, \quad ad - bc = (-1)^{\lambda+\mu+1} = 1,$$

where a, b, c and d are positive integers and hence (6.1) is a member of the group Γ . If we let

$$-\eta'_1 = [0, s_\lambda, s_{\lambda-1}, \dots, s_1, a_\mu, \dots, a_1, 1, a_{-1}, \dots, a_{-v}]$$

and

$$-\eta'_2 = [0, s_\lambda, s_{\lambda-1}, \dots, s_1, a_\mu, \dots, a_1, 1, a_{-1}, \dots, a_{-v} + 1],$$

then it is easily shown (this result is stated in Artin, *loc. cit.*, p. 173) that the relations,

$$\eta_1 = \frac{a\eta'_1 + b}{c\eta'_1 + d}, \quad \eta_2 = \frac{a\eta'_2 + b}{c\eta'_2 + d},$$

hold, where a, b, c and d are given in (6.1). Thus it is seen that there is a transformation of the group Γ_2 which transforms the rectangle

$$\bar{\Delta} = \{1, a_1, \dots, a_\mu, s_1, \dots, s_\lambda, 1; 0, a_{-1}, a_{-2}, \dots, a_{-v}\}$$

into the rectangle

$$\bar{\Delta}' = \{1; 0, s_\lambda, s_{\lambda-1}, \dots, s_1, a_\mu, \dots, a_1, 1, a_{-1}, \dots, a_{-v}\}.$$

Using only transformations of the group Γ_2 , each of the non-overlapping rectangles of σ_λ , for fixed λ , can be transformed into one of the non-overlapping set $\sigma'_\lambda, \{1; s_\lambda, \dots, s_1, a_\mu, \dots, a_1, 1, a_{-1}, \dots, a_{-v}\}$, such that the correspondence is one-to-one.

LEMMA 6.2. *There exists a positive constant k_4 such that*

$$\mu(\sigma'_\lambda) > k_4 \mu(\sigma_\lambda).$$

Considering a single pair, $\bar{\Delta}$ and $\bar{\Delta}'$, of corresponding rectangles of the sets σ_λ and σ'_λ ,

$$\begin{aligned} \mu(\bar{\Delta}) &= |\xi_2 - \xi_1| \cdot |\eta_2 - \eta_1|, \\ \text{and} \quad \mu(\bar{\Delta}') &= |\eta'_2 - \eta'_1|. \end{aligned}$$

A brief computation yields the equalities

$$\begin{aligned} \frac{\mu(\bar{\Delta}')}{\mu(\bar{\Delta})} &= \frac{|\eta'_2 - \eta'_1|}{|\xi_2 - \xi_1| \cdot |\eta_2 - \eta_1|} \\ &= |(c\xi'_1 + d)(c\xi'_2 + d)(c\eta'_1 + d)(c\eta'_2 + d)| = \frac{|(\xi'_1 - \eta'_1)(\xi'_2 - \eta'_2)|}{|(\xi_1 - \eta_1)(\xi_2 - \eta_2)|}. \end{aligned}$$

From the inequalities $1 \leq \xi_1, \xi'_1, \xi_2, \xi'_2 \leq 2$, $-1 \leq \eta_1, \eta'_1, \eta_2, \eta'_2 \leq 0$, it follows from the last equation that

$$\frac{\mu(\bar{\Delta}')}{\mu(\bar{\Delta})} \geq \frac{1}{9}.$$

But this inequality holds for each of the corresponding pairs in σ_λ and σ'_λ and the desired lemma holds with $k_4 = 1/9$.

LEMMA 6.3. *If E is a measurable point set of π which is invariant under the group Γ_2 , then*

$$\mu(E \cdot \sigma_\lambda) > k_4 \mu(E \cdot \sigma'_\lambda).$$

Denoting, as above, by $\bar{\Delta}$ and $\bar{\Delta}'$ corresponding rectangles of the sets σ_λ and σ'_λ respectively, since E is an invariant set, the set $E \cdot \bar{\Delta}$ is transformed into the set $E \cdot \bar{\Delta}'$ by the transformation of the group Γ_2 taking $\bar{\Delta}$ into $\bar{\Delta}'$. Using the equation

$$\frac{\mu(\bar{\Delta})}{\mu(\bar{\Delta}')} = \frac{|(\xi_1 - \eta_1)(\xi_2 - \eta_2)|}{|(\xi'_1 - \eta'_1)(\xi'_2 - \eta'_2)|}$$

there is obtained as above

$$\mu(\bar{\Delta}) \geq k_4 \mu(\bar{\Delta}').$$

But if $\bar{\bar{\Delta}}$ is any sub-rectangle, with sides parallel to the axes, of $\bar{\Delta}$, and $\bar{\bar{\Delta}}'$ is the corresponding rectangle under T , precisely the same inequality can be obtained, viz.,

$$\mu(\bar{\bar{\Delta}}) \geq k_4 \mu(\bar{\bar{\Delta}}').$$

The rectangle $\bar{\bar{\Delta}}$ being an arbitrary sub-rectangle (with sides parallel to the axes) of $\bar{\Delta}$, it follows that the Jacobian of T , evaluated at any point of $\bar{\Delta}$, lies between 9 and $1/9$. This implies the inequality $\mu(E \cdot \bar{\Delta}) \geq k_4 \mu(E \cdot \bar{\Delta}')$. By summing over the set σ_λ , the stated lemma is obtained.

LEMMA 6.4. *E being a measurable invariant set of π , there exists a positive constant, k_5 , which does not depend on how Δ is chosen in D , and is such that if λ is chosen sufficiently large,*

$$\mu(E \cdot \sigma'_\lambda) \geq k_5 \mu(E \cdot D) \mu(\sigma'_\lambda).$$

Case I. $E \cdot D = \{1, e_1, e_2, \dots, e_m; 0, e_{-1}, e_{-2}, \dots, e_{-n}\}$.

Let λ be so chosen that $\lambda > n$. Then

$$\mu(E \cdot \sigma'_\lambda) = L(e_1, \dots, e_m) \sum_{s_{\lambda-n} \dots s_1} L(e_{-1}, \dots, e_{-n}, s_{\lambda-n}, \dots, s_1, a_\mu, \dots, a_1, 1, a_{-1}, \dots, a_{-n}),$$

where $\sum_{s_{\lambda-n} \dots s_1}$ indicates the sum of the lengths of the intervals for which $s_{\lambda-n}, \dots, s_1$ are arbitrary positive integers, but the other elements are fixed.

Similarly,

$$\mu(\sigma'_\lambda) = \sum_{s_\lambda \dots s_1} L(s_\lambda, s_{\lambda-1}, \dots, s_1, a_\mu, \dots, a_1, 1, a_{-1}, \dots, a_{-n}),$$

where the sum is extended over all positive integral values of s_λ, \dots, s_1 .

From Lemma 5.1,

$$k_1 \frac{L(1, \dots, 1, a_\mu, \dots, a_{-v})}{L(1, \dots, 1)} < \frac{L(s_\lambda, s_{\lambda-1}, \dots, s_1, a_\mu, \dots, a_{-v})}{L(s_\lambda, \dots, s_1)} < k_2 \frac{L(1, \dots, 1, a_\mu, \dots, a_{-v})}{L(1, \dots, 1)}$$

where in each term where $1, \dots, 1$ occurs, there are λ such elements. It follows from this that

$$\sum_{s_\lambda \dots s_1} L(s_\lambda, \dots, s_1, a_\mu, \dots, a_{-v}) < k_2 \frac{L(1, \dots, 1, a_\mu, \dots, a_{-v})}{L(1, \dots, 1)} \sum_{s_\lambda \dots s_1} L(s_\lambda, \dots, s_1).$$

Evidently

$$\sum_{s_\lambda \dots s_1} L(s_\lambda, \dots, s_1) = 1,$$

and

$$\sum_{s_{\lambda-n} \dots s_1} L(e_{-1}, e_{-2}, \dots, e_{-n}, s_{\lambda-n}, \dots, s_1) = L(e_{-1}, e_{-2}, \dots, e_{-n}).$$

With the aid of these

$$\frac{\mu(E \cdot \sigma'_\lambda)}{\mu(\sigma'_\lambda)} \geq k_1 k_2^{-1} L(e_{-1}, e_{-2}, \dots, e_{-n}) L(e_1, \dots, e_m) = k_1 k_2^{-1} \mu(E \cdot D).$$

Choosing $k_5 = k_1/k_2$, the lemma holds in this case.

Case II. $E \cdot D$ is a finite set of non-overlapping rectangles of the net in D .

Let $E \cdot D = \sum_{i=1}^N R_i$. If λ is chosen sufficiently large the proof given in Case I holds simultaneously for all of the set R_i , ($i = 1, \dots, N$), and hence

$$\mu(E \cdot \sigma'_\lambda) = \sum_{i=1}^N \mu(R_i \cdot \sigma'_\lambda) \geq \sum_{i=1}^N k_1 k_2^{-1} \mu(R_i) \mu(\sigma'_\lambda) = k_1 k_2^{-1} \mu(\sigma'_\lambda) \mu(E \cdot D).$$

The lemma holds again with $k_5 = k_1 k_2^{-1}$.

Case III. $E \cdot D$ is an infinite set of non-overlapping rectangles of the net in D ; $E \cdot D = \sum_{i=1}^{\infty} R_i$.

Given ϵ , there exists an N such that $\mu(\sum_{i=1}^N R_i) > (1 - \epsilon) \mu(E \cdot D)$. For any λ , $\mu(E \cdot \sigma'_\lambda) = \sum_{i=1}^{\infty} \mu(R_i \cdot \sigma'_\lambda) \geq \sum_{i=1}^N \mu(R_i \cdot \sigma'_\lambda)$. If λ is sufficiently large, Case II can be applied and

$$\mu(E \cdot \sigma'_\lambda) \geq \sum_{i=1}^N \mu(R_i \cdot \sigma'_\lambda) \geq k_1 k_2^{-1} \mu(\sigma'_\lambda) \sum_{i=1}^N \mu(R_i) > k_1 k_2^{-1} (1 - \epsilon) \mu(E \cdot D) \mu(\sigma'_\lambda).$$

If $\epsilon < 1/3$, the lemma holds with $k_5 = 2k_1/3k_2$.

Case IV. $E \cdot D$ is an open set. This case is already included under III, for an open set is the sum of an infinite set of non-overlapping rectangles of the net in D , together with the boundaries of these rectangles. Since the boundaries form a zero set, they do not affect the argument.

Case V. $E \cdot D$ is a measurable set of positive measure.

Given $a_1, \dots, a_\mu, a_{-1}, \dots, a_{-v}$, from Lemma 5.1,

$$\begin{aligned} \mu(\sigma'_\lambda) &= \sum_{s_\lambda \dots s_1} L(s_\lambda, \dots, s_1, a_\mu, \dots, a_1, 1, a_{-1}, \dots, a_{-v}) \\ &\geq k_1 \frac{L(1, \dots, 1, a_\mu, \dots, a_{-v})}{L(1, \dots, 1)}, \end{aligned}$$

and hence $\mu(\sigma'_\lambda)$ is bounded away from zero, for arbitrary λ . Let $c > 0$ be such a lower bound.

Given $\epsilon = k_1 k_2^{-1} 3^{-1} c \mu(E \cdot D)$, there exists an open set E_0 such that $E_0 \supset E \cdot D$ and $\mu(E_0 - E \cdot D) < \epsilon$. For λ sufficiently large, we have from Case IV,

$$\mu(E_0 \cdot \sigma'_\lambda) > 2k_1 3^{-1} k_2^{-1} \mu(E_0) \mu(\sigma'_\lambda) \geq 2k_1 3^{-1} k^{-1} \mu(E \cdot D) \mu(\sigma'_\lambda).$$

But

$$\mu(E_0 \cdot \sigma'_\lambda) - \mu(E \cdot \sigma'_\lambda) \leq \mu(E_0 - E \cdot D) < \epsilon,$$

and hence

$$\mu(E \cdot \sigma'_\lambda) \geq 2k_1 3^{-1} k_2^{-1} \mu(E \cdot D) \mu(\sigma'_\lambda) - \epsilon.$$

From the choice of ϵ , it follows that

$$\mu(E \cdot \sigma'_\lambda) \geq k_1 3^{-1} k_2^{-1} \mu(E \cdot D) \mu(\sigma'_\lambda).$$

The desired lemma holds with $k_5 = k_1 3^{-1} k_2^{-1}$.

7. Metrical transitivity.

THEOREM 7.1. (*Metrical transitivity*). If E is a measurable set of π which is invariant under the group Γ_2 , either $\mu(E) = 0$ or $\mu[C_\pi(E)] = 0$.

It can be assumed that $\mu(E) > 0$. From Lemma 2.1, $\mu(E \cdot D) > 0$. Let Δ be a rectangle of the net in D . From Lemmas 6.1-6.4, if λ is chosen sufficiently large,

$$\begin{aligned} \mu(E \cdot \Delta) &\geq \mu(E \cdot \sigma'_\lambda) > k_4 \mu(E \cdot \sigma'_\lambda) \geq k_4 k_5 \mu(E \cdot D) \mu(\sigma'_\lambda) \\ &> k_4^2 k_5 \mu(E \cdot D) \mu(\sigma_\lambda) > k_3 k_4^2 k_5 \mu(E \cdot D) \mu(\Delta), \end{aligned}$$

or

$$\mu(E \cdot \Delta) > k \mu(\Delta),$$

where $k = k_3 k_4^2 k_5 \mu(E \cdot D) > 0$.

But this implies $\mu(E \cdot D) = \mu(D)$. For if this were not the case, there would be a point of D at which the metrical density of the set $E \cdot D$ would be zero. If P is such a point, a sufficiently small square, S , with P as center lies entirely in D and $\mu(E \cdot S) < k\mu(S)$. A sequence of non-overlapping rectangles of the net in D can be so chosen that $S = F + \sum_{i=1}^{\infty} R_i$, where F is a zero set. For each of these rectangles (7.2) holds and hence

$$\mu(E \cdot S) = \sum_{i=1}^{\infty} \mu(E \cdot R_i) > k \sum_{i=1}^{\infty} \mu(R_i) = k\mu(S).$$

This contradiction implies $\mu(E \cdot D) = \mu(D)$.

The set $C_{\pi}(E)$ is then a measurable invariant set such that $\mu(D \cdot C_{\pi}E) = 0$. From Lemma 2.1, $\mu(C_{\pi}E) = 0$, and Theorem 7.1 holds.

Theorem 7.1 implies the metrical transitivity of the group with respect to the real axis. For if this were not true there would exist a measurable non-zero set, E_1 , of the real axis such that E_1 would be invariant under the group Γ and $C(E_1)$, with respect to the real axis, would not be a linear zero set. But then the set $E_1 \times E_1$, consisting of those points (ξ, η) of π such that both ξ and η belong to E_1 , would be a measurable non-zero set of π , invariant under Γ_2 , such that neither $\mu(E_1 \times E_1) = 0$ nor $C_{\pi}(E_1 \times E_1) = 0$. This contradicts Theorem 7.1, and hence the group Γ must be metrically transitive with respect to the real axis.

Conversely, the group $^{\circ}G$ generated by the transformations

$$T^{(j)}: \quad \xi = x + a_j, \quad (j = 1, 2, \dots),$$

for which $\lim a_j = 0$, $n \rightarrow \infty$, $a_j \neq 0$, is metrically transitive with respect to the real axis, but the group G_2 is evidently not metrically transitive with respect to the plane.

BRYN MAWR COLLEGE,
BRYN MAWR, PA.

^{*} Martin, *loc. cit.*, p. 611.

SOME INTRINSIC AND DERIVED VECTORS IN A KAWAGUCHI SPACE.

By J. L. SYNGE.

1. *Introduction.* Let there be a space of N dimensions with coördinates x^i , and let there be a function F of

$$(1.1) \quad t, x^{(0)i}, x^{(1)i}, \dots, x^{(m)i},$$

where t is a parameter, and

$$(1.2) \quad x^{(q)i} = d^q x^i / dt^q, \quad x^{(0)i} = x^i.$$

We define F to be an invariant in the sense of tensor calculus. This does not imply the invariability of the functional form of F but simply that, when we employ a different coördinate system \bar{x}^i , we are to associate with it a function \bar{F} , given by the transformation of the arguments of F , t being unchanged: that is,

$$(1.3) \quad \bar{F}(t, \bar{x}^{(0)i}, \dots, \bar{x}^{(m)i}) = F(t, x^{(0)i}, \dots, x^{(m)i}).$$

The length of a curve $x^i = x^i(t)$ from $t = t_1$ to $t = t_2$ is defined to be the invariant

$$(1.4) \quad s = \int_{t_1}^{t_2} F dt.$$

Following Craig,¹ we shall call such a space a *Kawaguchi space of order m* , although Kawaguchi did not include t among the arguments of F .² If $m = 1$ in (1.1), if t is absent, and if F is homogeneous of the first degree in $x^{(1)i}$, the Kawaguchi space reduces to a Finsler space. If, more particularly, F is the square root of a homogeneous quadratic expression in $x^{(1)i}$, the Finsler space reduces to a Riemannian space. It will be noticed that we have made no reference to transformation of the parameter t . In both the Finsler space and the Riemannian space, the arc s as given by (1.4) is independent of the particular parameter t employed. In the Kawaguchi space, as discussed in the present paper, no condition is imposed on F to insure invariance of s under transformation of the parameter t . With the exception

¹ H. V. Craig, "On a generalized tangent vector," *American Journal of Mathematics*, vol. 57 (1935), p. 457.

² A. Kawaguchi, "Die Differentialgeometrie in der verallgemeinerten Mannigfaltigkeit," *Rendiconti Circolo Matematico di Palermo*, vol. 56 (1932), pp. 245-276.

of (5.30), the results established will be true if this invariance exists, but they do not require it.³

Since the parallel displacement of a vector, together with the associated ideas of absolute derivative and covariant derivative, play a fundamental part in Riemannian geometry, it is natural in studying these more general types of geometry to attempt to define parallel propagation and the associated operations in a way which reduces to the familiar definition when the space reduces to Riemannian space. The operation of absolute differentiation of a contravariant vector defined along a curve has been defined in the Finsler space by Taylor and Synge⁴ and in the Kawaguchi space of the second order by Craig,⁵ under the restriction stated. As I understand the work of Kawaguchi,² he appears to be interested in the most general forms which the operations in question could have, rather than the explicit development of the operations in terms of the function F . It is with this last development that the present paper is concerned.

Before proceeding to the discussion of absolute differentiation and parallel propagation along a curve, it is natural to develop the purely intrinsic properties of the curve itself. The vector which undergoes parallel propagation (unless *e. g.* it is a tangent vector) is not to be regarded as intrinsic.

For the Kawaguchi space of order m I develop a set of $m + 1$ intrinsic covariant vectors associated with a curve.⁵ When the space is Riemannian, $m = 1$, and there are just two of these vectors, corresponding to the tangent and first normal. In addition to these vectors, denoted by $\overset{p}{E}_i$, other intrinsic vectors are also developed. The mode of development is based on taking some invariant "generating function" H of the variables (1.1), or a larger set containing derivatives of higher orders with respect to the coördinates.

Passing on to the definition of the absolute derivative of a vector along a given curve, it appears that the most natural process is that which derives

³ In developing the theory of the Kawaguchi space for $m = 2$, H. V. Craig ("On parallel displacement in a non-Finsler space," *Transactions of the American Mathematical Society*, vol. 33 (1931), p. 129) subjects F to the condition that s shall be invariant under transformation of t . Craig's method has been extended to general values of m by H. Hombu, "On a non-Finsler metric space," *Tôhoku Mathematical Journal*, vol. 37 (1933), pp. 190-198.

⁴ J. H. Taylor, "A generalization of Levi-Civita's parallelism and the Frenet formulas," *Transactions of the American Mathematical Society*, vol. 27 (1925), pp. 246-264; J. L. Synge, "A generalization of the Riemannian line-element," *ibid.*, pp. 61-67. These papers were written simultaneously and independently.

⁵ Three of these vectors have been given by Craig: see ref. 1.

a covariant vector from a contravariant vector given along the curve. We are naturally most interested in formulae which involve only the *first* derivatives of the components of the given vector. We find that there are $m + 1$ such derived vectors for a Kawaguchi space of order m . One of these derived vectors may be picked out as the natural generalisation of the ordinary absolute derivative (whose vanishing implies parallel propagation), but it is interesting to note that the case $m = 1$ is rather peculiar as far as the general formula is concerned.

2. *The Eulerian vector E_i .* In connecting with the variational equation

$$(2.1) \quad \delta \int_{t_1}^{t_2} F dt = 0,$$

there are associated the well-known Eulerian equations. It is natural to expect that the expressions which are equated to zero in these equations are the components of a covariant vector. That is in fact the case, and it seems most natural to use a variational method to prove it. Craig¹ has established it by a direct method.

Let us take a singly infinite family of curves,

$$(2.2) \quad x^i = x^i(u, t),$$

where t is a variable parameter along each curve and u is constant along each curve. Then

$$(2.3) \quad \frac{d}{du} \int_{t_1}^{t_2} F dt = \int_{t_1}^{t_2} \frac{\partial F}{\partial u} dt,$$

and this is an invariant. Following Craig, we shall adopt the convenient notation

$$(2.4) \quad F_{(q)i} = \frac{\partial F}{\partial x^{(q)i}}, \quad F_{(q)i}^{..(p)} = \frac{d^p}{dt^p} \frac{\partial F}{\partial x^{(q)i}}.$$

In the present connection d/dt means partial differentiation with respect to t , u being held fixed. Then

$$(2.5) \quad \begin{aligned} \frac{\partial F}{\partial u} &= \sum_{q=0}^m F_{(q)i} \frac{\partial}{\partial u} x^{(q)i} \\ &= \sum_{q=1}^m F_{(q)i} \frac{d}{dt} \frac{\partial}{\partial u} x^{(q-1)i} + F_{(0)i} \frac{\partial x^i}{\partial u} \\ &= \frac{d}{dt} \sum_{q=1}^m F_{(q)i} \frac{\partial}{\partial u} x^{(q-1)i} + F_{(0)i} \frac{\partial x^i}{\partial u} - \sum_{q=1}^m F_{(q)i}^{..(1)} \frac{\partial}{\partial u} x^{(q-1)i}. \end{aligned}$$

Proceeding by successive steps in this way, we get

$$(2.6) \quad \frac{\partial F}{\partial u} = \frac{d}{dt} \left\{ \sum_{p=1}^m (-1)^{p-1} \sum_{q=p}^m F_{(q)i}^{..(p-1)} \frac{\partial}{\partial u} x^{(q-p)i} \right\} + E_i \frac{\partial x^i}{\partial u},$$

where

$$(2.7) \quad E_i = \sum_{q=0}^m (-1)^q F_{(q)i}^{(q)};$$

this is the Eulerian expression. We have then

$$(2.8) \quad \frac{d}{du} \int_{t_1}^{t_2} F dt = \left[\sum_{p=1}^m (-1)^{p-1} \sum_{q=p}^m F_{q(i)}^{(p-1)} \frac{\partial}{\partial u} x^{(q-p)i} \right]_{t=t_1}^{t=t_2} + \int_{t_1}^{t_2} E_i \frac{\partial x^i}{\partial u} dt.$$

Now let us suppose that the curves (2.2) are so chosen that the points for which $t = t_1$ and $t = t_2$ are common to them all, and further that the values of

$$x^{(1)i}, x^{(2)i}, \dots, x^{(m-1)i}$$

are also common to them all at these points. These conditions imply that

$$(2.9) \quad \frac{\partial}{\partial u} x^{(q)i} = 0 \text{ for } t = t_1 \text{ and } t = t_2, \quad (q = 0, 1, \dots, m-1).$$

Then the first term on the right-hand side of (2.8) vanishes, and we have

$$(2.10) \quad \frac{d}{du} \int_{t_1}^{t_2} F dt = \int_{t_1}^{t_2} E_i \frac{\partial x^i}{\partial u} dt.$$

Since this is an invariant, we have, on changing to new coördinates \bar{x}^i ,

$$(2.11) \quad \int_{t_1}^{t_2} \left(\bar{E}_i \frac{\partial \bar{x}^i}{\partial u} - E_j \frac{\partial x^j}{\partial u} \right) dt = 0,$$

or

$$(2.12) \quad \int_{t_1}^{t_2} \left(\bar{E}_i - E_j \frac{\partial x^j}{\partial \bar{x}^i} \right) \frac{\partial \bar{x}^i}{\partial u} dt = 0.$$

But $\partial \bar{x}^i / \partial u$ is arbitrary along any one of the curves $u = \text{const.}$, except at the end points, where its components vanish; hence, by the usual method of the calculus of variations,

$$(2.13) \quad \bar{E}_i = E_j \frac{\partial x^j}{\partial \bar{x}^i}.$$

Hence we have the following result:

THEOREM I. *The Eulerian expressions,*

$$(2.14) \quad E_i = \sum_{q=0}^m (-1)^q F_{(q)i}^{(q)},$$

are the components of a covariant vector in a Kawaguchi space of order m .

Although not intrinsic, we may mention an invariant, which appears as

a by-product of the preceding work. We have seen that the last term on the right of (2.8) is an invariant. Hence, if the terminal variations are left free, instead of being restricted by (2.9), it follows that the first term on the right of (2.8) is an invariant also. Let us put

$$(2.15) \quad X^i = \partial x^i / \partial u,$$

these being components of an arbitrary contravariant vector, given along any curve $u = \text{const.}$ The invariant in question is

$$(2.16) \quad \sum_{p=1}^m (-1)^{p-1} \sum_{q=p}^m F_{(q)i}^{(p-1)} (X^i)^{(q-p)}.$$

Writing it in a different form, we may state the following result:

THEOREM II. *The expression*

$$(2.17) \quad \sum_{r=0}^{m-1} (X^i)^{(r)} \sum_{p=1}^{m-r} (-1)^{p-1} F_{(p+r)i}^{(p-1)}$$

is an invariant in a Kawaguchi space of order m , X^i being the components of an arbitrary contravariant vector given along the curve $x^i = x^i(t)$.

When $m = 1$, this invariant is

$$(2.18) \quad F_{(1)i} X^i.$$

When $m = 2$, it is

$$(2.19) \quad X^i \{ F_{(1)i} - \frac{d}{dt} F_{(2)i} \} + \frac{dX^i}{dt} F_{(2)i}.$$

3. *The set of intrinsic vectors \bar{E}_i .* Since the establishment of the vector character of the Eulerian vector E_i , given in (2.14), involves nothing beyond the fact that F is an invariant function of the variables (1.1), it is obvious that we may obtain a class of vectors by the formula (2.14) on substituting for F any function $f(F)$ of it. In Riemannian geometry it is convenient to substitute F^2 . However, if the parameter t is chosen so as to make F constant along the curve, which can be done by taking for parameter

$$(3.1) \quad s = \int_{t_1}^t F dt,$$

it is easily seen that these vectors only differ from E_i by a constant factor.

It is to be borne in mind in all the subsequent work that new vectors may be obtained from those given by writing $f(F)$ instead of F .

Now let $\phi(t)$ be any function of t , transforming as an invariant on transformation of coördinates. Then ϕF is an invariant function of the

variables (1.1), and we may use it as a "generating function" instead of F . Substituting in (2.14), we deduce that

$$(3.2) \quad \sum_{q=0}^m (-1)^q (\phi F_{(q)i})^{(q)}$$

are the components of a vector: we have used the fact that ϕ involves t only. This reduces to

$$(3.3) \quad \sum_{q=0}^m (-1)^q \sum_{p=0}^q \binom{q}{p} \phi^{(p)} F_{(q)i}^{(q-p)},$$

or

$$(3.4) \quad \sum_{p=0}^m \phi^{(p)} \bar{E}_i^p,$$

where

$$(3.5) \quad \bar{E}_i^p = \sum_{q=p}^m (-1)^q \binom{q}{p} F_{(q)i}^{(q-p)}.$$

Here $\binom{q}{p}$ is the usual binomial symbol,

$$(3.6) \quad \binom{q}{p} = \frac{q(q-1) \cdots (q-p+1)}{p!}, \quad \binom{0}{0} = \binom{q}{0} = 1, \\ (p, q = 1, 2, \cdots; q \geq p).$$

Now for any assigned value of t , we may choose the values of $\phi, \phi^{(1)}, \cdots, \phi^{(m)}$ arbitrarily, and they are all invariants. Let us make them all zero except $\phi^{(r)}$, and let $\phi^{(r)} = 1$. Then the vector (3.4) reduces to \bar{E}_i^r , and we may state the following result:

THEOREM III. *The expressions*

$$(3.7) \quad \bar{E}_i^p = \sum_{q=p}^m (-1)^q \binom{q}{p} F_{(q)i}^{(q-p)}, \quad (p = 0, 1, \cdots, m),$$

are the components of a set of $m+1$ covariant vectors in a Kawaguchi space of order m .⁶

It may be of interest to write out a few of these vectors explicitly:

$$(3.8) \quad \left\{ \begin{array}{l} \bar{E}_i^0 = E_i = \sum_{q=0}^m (-1)^q F_{(q)i}^{(q)}, \\ \bar{E}_i^1 = \sum_{q=1}^m (-1)^q q F_{(q)i}^{(q-1)}, \\ \bar{E}_i^{m-2} = (-1)^m \{ F_{(m-2)i} - (m-1) F_{(m-1)i}^{(1)} + \frac{1}{2} m(m-1) F_{(m)i}^{(2)} \}, \\ \bar{E}_i^{m-1} = (-1)^{m-1} \{ F_{(m-1)i} - m F_{(m)i}^{(1)} \}, \\ \bar{E}_i^m = (-1)^m F_{(m)i}. \end{array} \right.$$

⁶ These vectors were obtained by Craig (ref. 1) for $p=0$, $p=1$, and $p=m$.

The vector character of the last is easily established directly.

For $m = 1$, there are only two vectors

$$(3.9) \quad \overset{0}{E}_i = F_{(0)i} - \frac{d}{dt} F_{(1)i}, \quad \overset{1}{E}_i = -F_{(1)i}.$$

For $m = 2$, there are only three vectors

$$(3.10) \quad \begin{cases} \overset{0}{E}_i = F_{(0)i} - \frac{d}{dt} F_{(1)i} + \frac{d^2}{dt^2} F_{(2)i}, \\ \overset{1}{E}_i = -F_{(1)i} + 2 \frac{d}{dt} F_{(2)i}, \\ \overset{2}{E}_i = F_{(2)i}. \end{cases}$$

4. *The intrinsic vectors* $\overset{p,r}{G}_i$. The quantities $x^{(1)i} = dx^i/dt$ are components of a contravariant vector, and hence any one of the expressions

$$(4.1) \quad x^{(1)j} \overset{p}{E}_j, \quad (p = 0, 1, \dots, m),$$

is an invariant. In deriving the vector character of $\overset{p}{E}_i$ as in (3.7), all that was required was the fact the F is an invariant function of the variables (1.1). Now (4.1) is an invariant, but it is a function of

$$(4.2) \quad t, x^{(0)i}, x^{(1)i}, \dots, x^{(2m-p)i},$$

since these are involved in $\overset{p}{E}_j$. Thus we may use (4.1) as a generating function instead of F in (3.7), provided that we extend the range of summation to include the variables (4.2). Hence we have the result:

THEOREM IV. *The expressions*

$$(4.3) \quad \overset{p,r}{G}_i = \sum_{q=r}^{2m-p} (-1)^q \binom{q}{r} (x^{(1)j} \overset{p}{E}_j) \overset{(q-r)}{(q)}_i, \\ (p = 0, 1, \dots, m; r = 0, 1, \dots, 2m - p),$$

where

$$(4.4) \quad \overset{p}{E}_j = \sum_{s=p}^m (-1)^s \binom{s}{p} \overset{(s-p)}{(s)}_j,$$

are the components of a set of $\frac{1}{2}(m+1)(3m+2)$ covariant vectors in a Kawaguchi space of order m .

Perhaps the most interesting of these vectors is that for which

$$(4.5) \quad p = m, \quad r = m - 1.$$

We have

$$(4.6) \quad G_i^{m,m-1} = (-1)^{m-1} \{ (x^{(1)j} E_j^m)_{(m-1)i} - m (x^{(1)j} E_j^m)_{(m)i}^{(1)} \} \\ = \delta_1^m F_{(1)i}^{(1)} - \delta_2^m F_{(2)i} - x^{(1)j} F_{(m)j(m-1)i} + m \frac{d}{dt} (x^{(1)j} F_{(m)j(m)i});$$

if $m = 1$, this becomes

$$(4.7) \quad G_i^{1,0} = F_{(1)i}^{(1)} - x^{(1)j} F_{(1)j(0)i} + \frac{d}{dt} (x^{(1)j} F_{(1)j(1)i}).$$

When the space is a Finsler space, F is homogeneous of degree unity in $x^{(1)i}$; then we have

$$(4.8) \quad G_i^{0,1} = F_{(1)i}^{(1)} - F_{(0)i},$$

which is the Eulerian vector, to within a sign. We may therefore state the following result:

THEOREM V. *In a Kawaguchi space of order m there is associated with each point of a curve a covariant vector $G_i^{m,m-1}$ given by (4.6). When the space is a Finsler space, this vector is identical with the Eulerian vector, except for sign.*

5. *The absolute derivative of a contravariant vector along a curve.* Let X^i be a contravariant vector field, the components being functions of the coördinates only. Let us take as generating function any one of the expressions

$$(5.1) \quad X^j E_j^p \quad (p = 0, 1, \dots, m).$$

This is a function of the variables (4.2), and hence may be substituted for F in (3.7), provided that the range of summation is suitably changed, the p of (3.7) being changed to another letter. In fact, we get a vector constructed as in (4.3). In order to show that the vectors obtained in this way are "derived" from X^j , we shall adopt the notation

$$(5.2) \quad D_{ij}^{p,r} X^j = \sum_{q=r}^{2m-p} (-1)^q \binom{q}{r} (X^j E_j^p)_{(q)i}^{(q-r)}, \\ (p = 0, 1, \dots, m; r = 0, 1, \dots, 2m - p).$$

We may state this result:

THEOREM VI. *The formula (5.2) defines a set of $\frac{1}{2}(m+1)(3m+2)$ covariant vectors derived (along a given curve) from a contravariant vector field X^i in a Kawaguchi space of order m , E_j^p being as given in (4.4).*

If $r = 0$, we shall have, in the course of the calculation, to take the partial derivative of X^j with respect to x^i . Consequently the derived vector will involve the partial derivatives of the vector field, as well as its derivatives with respect to t along the curve. But if we take $r > 0$, the formula will involve only the derivatives of X^j with respect to t , as in the formula for the absolute derivative in Riemannian space, which is

$$(5.3) \quad \frac{dX^i}{dt} + \left\{ \begin{matrix} i \\ jk \end{matrix} \right\} X^j \frac{dx^k}{dt}.$$

The formula (5.2) will involve the derivatives of X^j with respect to t up to and including the order $2m - p - r$. Let us confine our attention to those which involve derivatives of the first order only. We are then to take

$$(5.4) \quad 2m - p - r = 1, \quad r = 2m - p - 1.$$

Let us then consider the derived vectors

$$(5.5) \quad {}^{p, 2m-p-1} D_{ij} X^j = (-1)^{p+1} \{ (X^j \overset{p}{E}_j)_{(2m-p-1)t} - (2m-p) (X^j \overset{p}{E}_j)_{(2m-p)t}^{(1)} \}, \\ (p = 0, 1, \dots, m).$$

If

$$(5.6) \quad 2m - p - 1 = 0,$$

then $\partial X^j / \partial x^i$ will appear in the evaluation of this expression. Now (5.6) can be true only if $m = 1$, and then for the value $p = 1$. It is interesting to see what (5.5) gives in this particular case. We have

$$(5.7) \quad \overset{1,0}{D}_{ij} X^j = (X^j \overset{1}{E}_j)_{(0)t} - (X^j \overset{1}{E}_j)_{(1)t}^{(1)} \\ = \frac{\partial X^j}{\partial x^i} \overset{1}{E}_j + X^j (\overset{1}{E}_j)_{(0)t} - \frac{d}{dt} \{ X^j (\overset{1}{E}_j)_{(1)t} \},$$

where (cf. (3.8)) $\overset{1}{E}_j = -F_{(1)j}$: thus we have

$$(5.8) \quad \overset{1,0}{D}_{ij} X^j = -\frac{\partial X^j}{\partial x^i} F_{(1)j} - X^j F_{(1)j(0)t} + \frac{d}{dt} (X^j F_{(1)j(1)t}) \\ = \frac{dX^j}{dt} F_{(1)j(1)t} + X^j (F_{(1)j(1)t(1)k} x^{(2)k} \\ + F_{(1)j(1)t(0)k} x^{(1)k} - F_{(1)j(0)t}) - \frac{\partial X^j}{\partial x^i} F_{(1)j}.$$

To complete the case $m = 1$, we must also put $p = 0$: this gives

$$(5.9) \quad \overset{0,1}{D}_{ij} X^j = - (X^j \overset{0}{E}_j)_{(1)t} + 2 (X^j \overset{0}{E}_j)_{(2)t}^{(1)}.$$

Here we are to put (cf. (3.8))

$$(5.10) \quad \overset{0}{E}_j = F_{(0)j} - F_{(1)j}^{(1)} = F_{(0)j} - F_{(1)j(1)k}x^{(2)k} - F_{(1)j(0)k}x^{(1)k}.$$

Thus

$$\begin{aligned} (5.11) \quad \overset{0,1}{D}_{ij}X^j &= -X^j(\overset{0}{E}_j)_{(1)i} + 2\frac{d}{dt}\{X^j(\overset{0}{E}_j)_{(2)i}\} \\ &= -X^j(F_{(0)j(1)i} - F_{(1)j(1)k(1)i}x^{(2)k} \\ &\quad - F_{(1)j(0)i} - F_{(1)j(0)k(1)i}x^{(1)k}) - 2\frac{d}{dt}(X^jF_{(1)j(1)i}) \\ &= -2[F_{(1)j(1)i}\frac{dX^j}{dt} + X^j\{\frac{1}{2}(F_{(0)j(1)i} - F_{(1)j(0)i}) \\ &\quad + \frac{1}{2}F_{(1)j(1)i(0)k}x^{(1)k} + \frac{1}{2}F_{(1)j(1)i(1)k}x^{(2)k}\}]. \end{aligned}$$

We may state this result:

THEOREM VII. *In a Kawaguchi space⁷ of order 1 the formulae (5.8) and (5.11) define two covariant vectors derived from a contravariant vector field X^j : (5.8) involves the partial derivatives of X^j , but (5.11) involves only the derivatives with respect to t .*

It is interesting to see what the derived vectors (5.8) and (5.11) degenerate to in the case of a Finsler space. We shall not, however, use these formulae as they stand, but the corresponding formula with F^2 substituted for F . We shall denote the corresponding vectors by

$$(5.12) \quad \overset{1,0}{\Delta}_{ij}X^j, \quad \overset{0,1}{\Delta}_{ij}X^j.$$

We shall write

$$(5.13) \quad f_{ij} = \frac{1}{2}(F^2)_{(1)i(1)j}, \quad f_{ijk} = \frac{1}{2}(F^2)_{(1)i(1)j(1)k}.$$

We know that these are covariant tensors. Using the fact that F^2 is homogeneous of degree two in $x^{(1)i}$, we obtain

$$\begin{aligned} (5.14) \quad \overset{1,0}{\Delta}_{ij}X^j &= 2\left\{f_{ij}\frac{dX^j}{dt} + f_{ijk}X^jx^{(2)k}\right. \\ &\quad \left.+ \left(\frac{\partial f_{ij}}{\partial x^k} - \frac{\partial f_{jk}}{\partial x^i}\right)X^jx^{(1)k} - f_{jk}\frac{\partial X^j}{\partial x^i}x^{(1)k}\right\} \\ &= 2\left\{\left(f_{ij}\frac{\partial X^j}{\partial x^k} + X^j\frac{\partial f_{ij}}{\partial x^k}\right)\right. \\ &\quad \left.- \left(f_{jk}\frac{\partial X^j}{\partial x^i} + X^j\frac{\partial f_{jk}}{\partial x^i}\right)\right\}x^{(1)k} + 2f_{ijk}X^jx^{(2)k}. \end{aligned}$$

⁷ In the sense of the present paper, a Kawaguchi space of order 1 is not necessarily a Finsler space: for the Finsler space, F is homogeneous of degree unity in $x^{(1)i}$.

In the still more particular case of Riemannian space, we have

$$(5.15) \quad f_{ij} = g_{ij},$$

which are functions of the coördinates only. Then we have

$$(5.16) \quad \overset{1,0}{\Delta}_{ij} X^j = 2 \left(\frac{\partial X_i}{\partial x^k} - \frac{\partial X_k}{\partial x^i} \right) x^{(1)k},$$

whose vector character is well known.

For the other derived vector in a Finsler space, we have

$$(5.17) \quad -\frac{0,1}{2} \Delta_{ij} X^j = 2f_{ij} \frac{dX^j}{dt} + X^j \left\{ \left(\frac{\partial f_{ik}}{\partial x^j} - \frac{\partial f_{jk}}{\partial x^i} \right) x^{(1)k} + \frac{\partial f_{ij}}{\partial x^k} x^{(1)k} + f_{ijk} x^{(2)k} \right\},$$

$$(5.18) \quad \overset{0,1}{\Delta}_{ij} X^j = -4 \left\{ f_{ij} \frac{dX^j}{dt} + \frac{1}{2} \left(\frac{\partial f_{ij}}{\partial x^k} + \frac{\partial f_{ik}}{\partial x^j} - \frac{\partial f_{jk}}{\partial x^i} \right) X^j x^{(1)k} + \frac{1}{2} f_{ijk} X^j x^{(2)k} \right\}.$$

Hence we may state this result:

THEOREM VIII. *When the Kawaguchi space of order 1 is a Finsler space the covariant vectors (5.8) and (5.11) derived from a contravariant vector field X^j degenerate (when F is replaced by F^2) to (5.14) and (5.18).*

Except for the numerical factor -4 , formula (5.18) expresses the covariant components of the absolute derivative of a contravariant vector in a Finsler space, as defined by Taylor and Synge.⁴ Since the vanishing of this absolute derivative is taken to define parallel propagation in a Finsler space, it seems appropriate to adopt the following definition of parallel propagation in a Kawaguchi space of order 1, which, it is to be remembered, differs from a Finsler space only in so far as F is not necessarily homogeneous of order one in $x^{(1)i}$.

DEFINITION. *A vector X^j is propagated parallelly along a curve in a Kawaguchi space of order 1 when it satisfies the equations*

$$(5.19) \quad \overset{0,1}{\Delta}_{ij} X^j = 0,$$

where

$$(5.20) \quad -\frac{0,1}{2} \Delta_{ij} X^j = (F^2)_{(1)i(2)j} \frac{dX^j}{dt} + \frac{1}{2} \{ (F^2)_{(1)i(0)j} - (F^2)_{(1)j(0)i} \} X^j + \frac{1}{2} (F^2)_{(1)i(1)j(0)k} X^j x^{(1)k} + \frac{1}{2} (F^2)_{(1)i(1)j(1)k} X^j x^{(2)k}.$$

Let us now return to the general Kawaguchi space of order m and the derived vectors given in (5.5). Let us put

$$(5.21) \quad p = m - 1,$$

so that the formula reads

$$(5.22) \quad {}^{m-1,m}D_{ij} X^j = (-1)^m \{ (X^j E_j)_{(m)i} - (m+1) (X^j E_j)_{(m+1)i}^{(1)} \}.$$

We have already investigated this for $m = 1$ (cf. (5.9)) and we have seen that, if we put F^2 in place of F , it reduces to a familiar expression in a Finsler space. Let us now reduce the expression for any value of m . We have

$$(5.23) \quad {}^{m-1,m}D_{ij} X^j = (-1)^m [X^j (E_j)_{(m)i} - (m+1) \frac{d}{dt} \{X^j (E_j)_{(m+1)i}\}].$$

Let us turn to (3.8); we have

$$(5.24) \quad \begin{aligned} E_j &= (-1)^{m-1} (F_{(m-1)j} - m F_{(m)j}^{(1)}) \\ &= (-1)^{m-1} (F_{(m-1)j} - m \sum_{q=0}^m F_{(m)j(q)k} x^{(q+1)k}), \end{aligned}$$

and so

$$(5.25) \quad \begin{aligned} (E_j)_{(m)i} &= (-1)^{m-1} (F_{(m-1)j(m)i} - m F_{(m)j(m-1)i} - m \sum_{q=0}^m F_{(m)j(q)k(m)i} x^{(q+1)k}), \\ (E_j)_{(m+1)i} &= -(-1)^{m-1} m F_{(m)j(m)i}. \end{aligned}$$

Thus (5.23) reads

$$(5.26) \quad \begin{aligned} {}^{m-1,m}D_{ij} X^j &= -m(m+1) F_{(m)i(m)j} \frac{dX^j}{dt} \\ &\quad + (m F_{(m-1)i(m)j} - F_{(m)i(m-1)j} - m^2 \sum_{q=0}^m F_{(m)i(m)j(q)k} x^{(q+1)k}) X^j, \end{aligned}$$

which agrees with (5.11) when we put $m = 1$.

We get a derived vector analogous to (5.26) if we substitute for F any function of F . The earlier work indicates that F^2 is the most suitable function to take. We may state the following result:

THEOREM IX. *In a Kawaguchi space of order m the formula*

$$(5.27) \quad \begin{aligned} \Delta_{ij}^{m-1,m} X^j &= -m(m+1) (F^2)_{(m)i(m)j} \frac{dX^j}{dt} \\ &\quad + \{ m (F^2)_{(m-1)i(m)j} - (F^2)_{(m)i(m-1)j} \\ &\quad - m^2 \sum_{q=0}^m (F^2)_{(m)i(m)j(q)k} x^{(q+1)k} \} X^j \end{aligned}$$

defines a covariant vector along any curve $x^i = x^i(t)$ along which the contravariant vector X^j is given.

The following definitions may be set down:

DEFINITION. $\overset{m-1,m}{\Delta}_{ij} X^j$ is the absolute (covariant) derivative of the vector X^i .

DEFINITION. A vector X^j is propagated parallelly along a curve in a Kawaguchi space of order m if its components satisfy the differential equations

$$(5.28) \quad \overset{m-1,m}{\Delta}_{ij} X^j = 0.$$

The type of geometry with which the present paper deals is essentially concerned with processes of generalisation. Generalisations are by no means unique, and a method such as that developed in the present paper opens up an embarrassing variety of generalisations. Riemannian geometry is the well-established base with which (as a particular case) our generalisations are to be checked. The definitions adopted above do check with well-established results in Riemannian space, but we may ask whether, in the case of the Kawaguchi space of order m , we have been wise to use F^2 as the generating function in (5.27), instead of F^{m+1} . This would equally well give agreement with results in Riemannian or Finsler space, since when $m = 1$ we have $m + 1 = 2$.

It is easily seen that

$$(5.29) \quad f_{ij} = \frac{1}{2} (F^2)_{(m)i(m)j}$$

is a covariant tensor in a Kawaguchi space of order m ; if the determinant of f_{ij} is not zero, we may introduce a conjugate contravariant tensor f^{ij} , and use it to convert covariant vectors into contravariant vectors. Thus although the formula (5.27) derives a covariant vector from a contravariant vector X^j , we can at once obtain a contravariant derived vector,

$$(5.30) \quad \overset{m-1,m}{\Delta}{}^{k,j} X^j = f^{k,i} \overset{m-1,m}{\Delta}_{ij} X^j.$$

It should be noted, however, that this cannot be done if s is invariant under transformation of t , for in that case the determinant of f_{ij} vanishes.^s

UNIVERSITY OF TORONTO.

^s Cf. H. V. Craig, *Bulletin of the American Mathematical Society*, vol. 36 (1930), p. 560.

AN ANALYTIC CHARACTERIZATION OF SURFACES OF FINITE LEBESGUE AREA.¹ PART I.

By CHARLES B. MORREY, JR.²

Since Schwarz³ showed that the ordinary definition of the length of a curve could not be generalized directly to give a definition of the area of a surface, many definitions of the area of a surface have been proposed. In this paper, we shall use that proposed by Lebesgue in his thesis.⁴ Although this definition was almost forgotten for over twenty years due to the lack of methods for handling it and also perhaps for esthetic reasons, its usefulness in connection with the solutions of the Problem of Plateau (particularly those of Radó⁵ and McShane⁶) demonstrates its value as a tool in Analysis and Geometry.

This definition presupposes a definition of limit elements in the field of surfaces. For surfaces $z = f(x, y)$, it is clear that we should say that a sequence of surfaces S_n had the surface S as its limit if and only if the corresponding functions $f_n(x, y)$ converged uniformly to the limiting $f(x, y)$. An ideal extension of this definition to general surfaces is furnished by Fréchet's definition of the distance between two surfaces.⁷ It is a curious fact that, although earlier workers in the area of surfaces (such as Lebesgue and Geöcze) clearly had some such definition of convergent sequences in mind, it was not precisely formulated until so recently and has accordingly been used only in the work of Radó, McShane, Douglas, and the author.

The problem considered in this paper is that of determining an analytic

¹ Part I was presented to the American Mathematical Society on December 27, 1932, under the title "An analytic criterion that a surface possess finite Lebesgue area." Part II was presented on April 14, 1933, under its present title.

² National Research Fellow (1931-33).

³ H. A. Schwarz, *Gesammelte Anhandlungen*, vol. 1, p. 309.

⁴ H. Lebesgue, "Intégral, longueur, aire" (Dissertation), *Annali di Matematica*, ser. 3, vol. 7 (1902), pp. 231-359.

⁵ T. Radó, "On the problem of least area and the problem of Plateau," *Mathematische Zeitschrift*, vol. 32 (1930), pp. 763-796.

⁶ E. J. McShane, "Parametrizations of saddle surfaces with application to the problem of Plateau," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 716-733.

⁷ M. Fréchet, "Sur la distance de deux surfaces," *Annales de la Société Polonaise de Mathématiques*, vol. 3 (1924), pp. 4-19.

characterization of surfaces of finite area more or less analogous to that of rectifiable curves. Accordingly, we shall list mainly researches on this problem and refer the reader to Radó's *Ergebnisse* tract "On the Problem of Plateau" ⁸ for the most important literature on the general theory of the area. Surfaces $z = f(x, y)$ of finite Lebesgue area were first characterized by Geöcze ⁹ and later independently by Tonelli. ¹⁰ Tonelli also characterized functions $f(x, y)$ (calling them absolutely continuous) for which the area of the surface $z = f(x, y)$ is given by the classical integral formula, and these functions have been invaluable in subsequent work on area. McShane ¹¹ and the author ¹² independently defined a class of representations (called class L) of surfaces for which $L(S)$ is finite and given by the classical integral formula and McShane ¹³ characterized "saddle" surfaces of finite area bounded by Jordan curves by showing that each such surface possesses a representation of class L (in fact "generalized conformal").

The present paper gives an analytic characterization of the most general surface of finite Lebesgue area. It is first shown (in part I) that every non-degenerate (see § 1) surface of finite Lebesgue area possesses a generalized conformal (see § 2) representation. To characterize arbitrary surfaces, it is found helpful to allow parametric representations of surfaces on certain sets in 3-space called *hemicactoids*, a theory of such representations having been fully developed in the author's recent paper "The topology of (path) surfaces" ¹⁴ which will hereafter be referred to as T. It is then shown in § 3 (part II) that a necessary and sufficient condition for a surface S to possess finite Lebesgue area is that there exists a hemicactoid \bar{H} on which S may be represented, the representation being generalized conformal on each non-degenerate cyclic element (see § 3).

Throughout this paper we shall use the following vector notation: the

⁸ T. Radó, "On the Problem of Plateau," *Ergebnisse der Mathematik und Ihrer Grenzgebiete* (Springer), vol. 2 (1933).

⁹ Z. de Geöcze, "Die notwendigen und hinreichenden Bedingungen für einer endlichen Flächeninhalt eines Flächenstückes," *Mathematicai es Fizikai Lapok*, vol. 25 (1916), pp. 61-81.

¹⁰ L. Tonelli, "Sulla quadratura delle superficie," *Atti della Reale Accademia dei Lincei*, ser. 6, vol. 3 (1926), pp. 357-362, 445-450, 633-638, 714-719.

¹¹ E. J. McShane, "Integrals over surfaces in parametric form," *Annals of Mathematics*, vol. 34 (1933), pp. 815-838.

¹² C. B. Morrey, Jr., "A class of representations of manifolds (Part I)," *American Journal of Mathematics*, vol. 55 (1933), pp. 683-707 (hereafter cited as R).

¹³ *Loc. cit.* (first reference).

¹⁴ C. B. Morrey, Jr., "The topology of (path) surfaces," *American Journal of Mathematics*, vol. 57, no. 1 (January, 1935), pp. 17-50.

letters x and X shall stand for the coördinates (x^1, \dots, x^N) and (X^1, \dots, X^N) of a point in the x -space in which the given surface lies, the letters u and U for (u, v) and (U, V) respectively, the sum and difference of pairs of these letters, i. e., $x_1 \pm x_2$ or $u_1 \pm u_2$, will denote the vector sum and differences in the respective spaces, x_α will stand for the vector $(\partial x^1/\partial \alpha, \dots, \partial x^N/\partial \alpha)$, α being a parameter, $x(u)$ and $X(U)$ will be vector functions, and if ϕ is a vector in any space, $|\phi|$ shall denote its length. We shall sometimes write $x(P)$ to mean $x(u)$ where u is the coördinate vector of the point P . Given a point set E , \bar{E} shall denote its closure and E^* the set of its frontier points. The letters r and R shall always denote Jordan regions (i. e., regions bounded by a single Jordan curve). All vector functions occurring in a transformation or a representation of a surface will be assumed to be continuous.

1. *Non-degenerate vector functions.* In this section, we shall merely demonstrate a few simple properties of such vector functions which, however, are invaluable in the developments of the next section.

Definition 1. Let $x(u)$ be a (continuous vector) function defined on \bar{r} . We define the *oscillation* of $x(u)$ over the set, E , as the least upper bound of $|x(u) - x(u')|$, for all u, u' in E . (T, def. 1, § 3.)

Definition 2. Let $x(u)$ be defined on \bar{r} and suppose C is a continuum, in \bar{r} , of diameter $\geq \rho > 0$. We define $\eta_1(\rho, x; C)$ as the oscillation of $x(u)$ over C and $\eta_1(\rho, x)$ the greatest lower bound of $\eta_1(\rho, x; C)$ for all such C .

Definition 3. We shall say that a continuum, C , is the *upper limit* of a sequence, $\{C_n\}$, of continua if (T, def. 1, § 2)

- (i) all the limit points of a sequence, $\{P_n\}$, of points, $P_n \in C_n$, lie on C ;
- (ii) if P is any point of C , there is a sequence, $\{P_k\}$, of points, $P_k \in C_{n_k}$, which converges to P , $\{n_k\}$ being a subsequence of the integers.

If C is also the upper limit of every subsequence of $\{C_n\}$, then we say that C is the *limit* of the sequence $\{C_n\}$ and that $\{C_n\}$ *converges* to C .

- The following lemma is well known:¹⁵

LEMMA 1. If $\{C_n\}$ is a sequence of continua in a closed bounded region \bar{R} , then a subsequence of $\{C_n\}$ possesses a unique limit continuum, C . Thus the sets (1) of all continua of \bar{R} , and (2) of all continua in \bar{R} of diameter $\geq \rho$, are compact.

¹⁵ See, for instance, R. L. Moore, "Foundations of point set theory," *American Mathematical Society Colloquium Publications*, vol. 13, pages 28, 29.

THEOREM 1. Suppose $x(u)$ is defined and continuous on \bar{r} . Then $\eta_1(\rho, x; C)$ is lower semicontinuous in C , and thus takes on its minimum, $\eta_1(\rho, x)$, on some continuum \bar{C} of diameter $\geq \rho$.

Proof. Let $\{C_n\}$ be a sequence of continua, of diameter $\geq \rho$, with limit continuum C . Let P_1 and P_2 be points of C such that

$$\eta_1(\rho, x; C) = |x(P_1) - x(P_2)|.$$

We may select a subsequence $\{n_k\}$ of the positive integers such that $P_i^{n_k} \rightarrow P_i$ ($P_i^{n_k} \in C_{n_k}$) ($i = 1, 2$), and $\eta_1(\rho, x; C_{n_k}) \rightarrow \varliminf_{n \rightarrow \infty} \eta_1(\rho, x; C_n)$. Then clearly

$$\begin{aligned} \eta_1(\rho, x; C) &= |x(P_1) - x(P_2)| = \lim_{k \rightarrow \infty} |x(P_1^{n_k}) - x(P_2^{n_k})| \\ &\leq \lim_{k \rightarrow \infty} \eta_1(\rho, x; C_{n_k}), \end{aligned}$$

which proves the theorem.

THEOREM 2. If the (continuous vector) functions $x_n(u)$, defined on \bar{r} , approach $x(u)$ uniformly,

$$\eta_1(\rho, x) \leq \varliminf_{n \rightarrow \infty} \eta_1(\rho, x_n).$$

Proof. For each n , there exists a continuum C_n , of diameter $\geq \rho$, such that $\eta_1(\rho, x_n) = \eta_1(\rho, x_n; C_n)$. We may select a subsequence, $\{n_k\}$, of integers such that $C_{n_k} \rightarrow C$ and $\eta_1(\rho, x_{n_k}) \rightarrow \varliminf_{n \rightarrow \infty} \eta_1(\rho, x_n)$. Now let P_1 and P_2 be points on C for which $|x(P_1) - x(P_2)|$ is a maximum, and let $\{n_i\}$ be a subsequence of the integers $\{n_k\}$ so that we can find points, P_{i, n_i} , on C_{n_i} , so that $P_{i, n_i} \rightarrow P_i$, ($i = 1, 2$). Then clearly

$$\begin{aligned} \eta_1(\rho, x) &\leq |x(P_1) - x(P_2)| = \lim_{i \rightarrow \infty} |x_{n_i}(P_{1, n_i}) - x_{n_i}(P_{2, n_i})| \\ &\leq \lim_{i \rightarrow \infty} \eta_1(\rho, x_{n_i}) = \varliminf_{n \rightarrow \infty} \eta_1(\rho, x_n), \end{aligned}$$

which proves the theorem.

Definition 4. A vector function is said to be *non-degenerate* on a continuum, \bar{C} , if it is not constant over any continuum of \bar{C} containing more than one point (cf. T, def. 5, § 4).

The following two theorems follow immediately from the definitions.

THEOREM 3. If $x(u)$ is non-degenerate on \bar{r} , $\eta_1(\rho, x) > 0$ if $\rho > 0$.

THEOREM 4. If $x(u)$ is non-degenerate on \bar{r} , $u = u(U)$ is a 1 — 1

continuous transformation of \bar{r} into \bar{R} , and we define $X(U) = x[u(U)]$, then $X(U)$ is non-degenerate on \bar{R} .

The following theorem simplifies the argument in § 2:

THEOREM 5. *If $\{x_n(u)\}$ is a sequence of non-degenerate vector functions approaching the non-degenerate vector function $x(u)$ uniformly, we can find a function $\eta_1(\rho)$, positive for $\rho > 0$, such that*

$$\eta_1(\rho) \leq \eta_1(\rho, x); \quad \eta_1(\rho) \leq \eta_1(\rho, x_n), \quad (n = 1, 2, \dots).$$

Proof. We may define $\eta_1(\rho)$ as the greatest lower bound of $\eta_1(\rho, x)$ and the numbers $\eta_1(\rho, x_n)$. If this is zero for some $\rho > 0$, we may extract a subsequence, $\{n_k\}$, of the positive integers so that $\eta_1(\rho, x_{n_k}) \rightarrow 0$ which contradicts Theorems 2 and 3.

2. *The existence of a generalized conformal representation of an arbitrary non-degenerate surface of finite Lebesgue area.* In this section, we prove a selection theorem for a sequence of representations of non-degenerate surfaces which converge to a non-degenerate surface. This theorem together with its proof is the exact analog for the vector functions representing these surfaces of Lebesgue's selection theorem for a sequence of monotone functions with uniformly bounded Dirichlet integrals which converges uniformly on the boundary of a region. By means of this theorem, the main result of the paper is established. The method of proof used extends to representations on an n -sphere of n -dimensional manifolds which are of class L with

$$\int_{R_n} \cdots \int [g_{11} + \cdots + g_{nn}]^{n/2} du^1 \cdots du^n < M$$

independent of n , the g_{ii} being among the coefficients g_{ij} of the fundamental (positive definite) form

$$ds^2 = \sum_{i=1}^n \sum_{j=1}^n g_{ij} du^i du^j.$$

Definition 1. A function, $f(x, y)$, will be said to be *absolutely continuous in the sense of Tonelli*¹⁶ (A. C. T.) in a region \bar{r} if it is A. C. T. on every rectangle interior to r with f_x and f_y summable over r ; $f(x, y)$ is A. C. T.¹⁷ for $a \leq x \leq b$, $c \leq y \leq d$ if it is continuous there and

¹⁶ L. Tonelli, "Sulla quadratura delle superficie," *Atti della Reale Accademia Nazionale dei Lincei*, ser. 6, vol. 3 (1926), pp. 633-638.

¹⁷ In a paper, "Complements of potential theory II," *American Journal of Mathematics*, vol. 55 (1933), pp. 42-46, G. C. Evans has shown that this concept is identical with that of a continuous potential function of its generalized derivatives (see ref. 19).

(i) for almost every X , $a \leq X \leq b$, $f(X, y)$ is absolutely continuous in y , and for almost every Y , $c \leq Y \leq d$, $f(x, Y)$ is absolutely continuous in x ;

(ii) $\int_a^b V_c^{(y)d}[f(X, y)]dX$ and $\int_c^d V_a^{(x)b}[f(x, Y)]dY$ both exist, where $V_c^{(y)d}[f(X, y)]$, for instance denotes the variation of $f(X, y)$ on (c, d) considered as a function of y alone.

It is known that f_x and f_y exist almost everywhere and are summable.

The following definitions and lemmas may be found in the literature and are included here merely for the sake of completeness.

LEMMA 1.¹⁸ If $\{f_n(x, y)\}$ is a sequence of functions, A. C. T. on \bar{r} , which approach the continuous function, $f(x, y)$, uniformly and there exist constants M , $p > 1$, and $q > 1$, independent of n such that

$$\int_r \int [|\partial f_n / \partial x|^p + |\partial f_n / \partial y|^q] dx dy < M, \quad (n = 1, 2, \dots),$$

then $f(x, y)$ is A. C. T. on \bar{r} , $|f_x|^p$ and $|f_y|^q$ are summable on r , and

$$\begin{aligned} \int_r \int |f_x|^p dx dy &\leq \lim_{n \rightarrow \infty} \int_r \int |\partial f_n / \partial x|^p dx dy; \\ \int_r \int |f_y|^q dx dy &\leq \lim_{n \rightarrow \infty} \int_r \int |\partial f_n / \partial y|^q dx dy. \end{aligned}$$

LEMMA 2.¹⁹ Let $f(x, y)$ be A. C. T. in \bar{r}_1 , and let $x = x(s, t)$, $y = y(s, t)$ be a 1—1 transformation of \bar{r}_1 into \bar{R}_1 where $x(s, t)$ and $y(s, t)$ are continuous together with their first partial derivatives and $|x_s y_t - x_t y_s| \geq \lambda > 0$. Then if $\phi(s, t) = f[x(s, t), y(s, t)]$, we have that $\phi(s, t)$ is A. C. T. in \bar{R}_1 and

$$(2.1) \quad \phi_s = f_x x_s + f_y y_s, \quad \phi_t = f_x x_t + f_y y_t$$

almost everywhere.

LEMMA 3. Suppose (i) $f(x, y)$ is A. C. T. in \bar{r} with f_x^2 and f_y^2 summable over r , (ii) $x = x(s, t)$, $y = y(s, t)$ is a 1—1 conformal transformation (merely continuous on r^*) of \bar{r} into \bar{R} , (iii) $\phi(s, t) = f[x(s, t), y(s, t)]$. Then (a) $\phi(s, t)$ is A. C. T. on \bar{R} , (b) its partial derivatives are given almost everywhere by the formulas (2.1), (c) ϕ_s^2 and ϕ_t^2 are summable over R , and (d) we have

¹⁸ C. B. Morrey, Jr., *loc. cit.* (R).

¹⁹ G. C. Evans, "Fundamental points of potential theory," *Rice Institute Pamphlets*, vol. 7, no. 4 (1920), pp. 274-285, particularly.

$$\int_R (\phi_s^2 + \phi_t^2) ds dt = \int_r (f_x^2 + f_y^2) dx dy.$$

Proof. Lemma 3 is an immediate consequence of the preceding lemma as is easily seen by considering the mapping of regions entirely interior to R on regions entirely interior to r by the given transformation.

Definition 2.^{20, 21} A representation, $x = x(u)$, $u \in \bar{r}$, of a surface, S , is said to be of class L if

(i) the components, $x^i(u, v)$, are all A. C. T. on \bar{r} , ($i = 1, \dots, N$),

(ii) $\lim_{h \rightarrow 0} \int_{r_\alpha} \int \left| \frac{\partial(x_h^i, x_h^j)}{\partial(u, v)} - \frac{\partial(x^i, x^j)}{\partial(u, v)} \right| du dv = 0, \quad \alpha_0 > \alpha > 0,$
 $(i, j = 1, \dots, N),$

$$x_h^i(u, v) = (1/h^2) \int_u^{u+h} \int_v^{v+h} x^i(\xi, \eta) d\xi d\eta,$$

r_α being the set of points of r at a distance $\geq \alpha$ from r^* (i. e., this is true for all these α).

LEMMA 4.^{20, 21} A convenient subclass of representation of class L is determined by the following conditions:

(i) $x^i(u, v)$ is A. C. T., ($i = 1, \dots, N$),

(ii) $|x_u^i|^p, |x_v^j|^q$ are summable over r , $p, q \geq 1, 1/p + 1/q \leq 1$, ($i = 1, \dots, N$).

We include the case where one of p and q is unity and the other infinite by interpreting (ii), in the case where $p = 1, q = \infty$, for instance, to mean

(ii') $|x_v^i| < M, |x_u^i|$ summable in r , ($i = 1, \dots, N$).

Surfaces $z = f(x, y)$ with $f(x, y)$ A. C. T. are also seen to be of class L .

LEMMA 5.^{22, 23} If the representation, $x = x(u)$, of the surface S , is of class L , $L(S)$ is given by the usual integral formula.

*Definition 3.*²² The representation, $x = x(u)$, of the surface S is *generalized conformal* if it satisfies conditions (i) and (ii) of Lemma 4 with

²⁰ C. B. Morrey, Jr., *loc. cit.* (R).

²¹ E. J. McShane, *loc. cit.* (2nd ref., footnote 11).

²² C. B. Morrey, Jr., *loc. cit.* (R).

²³ E. J. McShane, *loc. cit.* (2nd ref.).

$p = q = 2$ and $E = G$, $F = 0$ almost everywhere, E , F , G being given by their usual formulas.

Definition 4. We say that the points P_1 and P_2 of a surface S , $S: x = x(u)$, $u \in \bar{r}$, are *logically distinct* if they correspond to distinct values u_1 and u_2 in \bar{r} , such that $x(u)$ is not constant over any continuum containing them both. This property is clearly invariant under changes of parameter, u . If S^{24} and $x(u)$ are non-degenerate, the above merely requires that $u_1 \neq u_2$.

LEMMA 6.²⁵ *Let Π be a non-degenerate polyhedron. It possesses a generalized conformal representation on the unit circle in which three given logically distinct points on the boundary of Π correspond to three given distinct points on the boundary of the unit circle. If Π is degenerate, the mapping is impossible.*

LEMMA 7.²⁶ *Let S , $S: x = x(u)$, and S_n , $S_n: x = x_n(u)$, ($n = 1, 2, \dots$), be continuous surfaces. Suppose (i) the given representations of the S_n are generalized conformal, (ii) the functions $x_n(u)$ converge uniformly to $x(u)$, and (iii) $\lim_{n \rightarrow \infty} L(S_n) = L(S)$. Then the given representation of S is generalized conformal.*

THEOREM 1. *Given that S , $S: x = x(u)$, $u \in \bar{r}$, and S_n , $S_n: x = x_n(u)$, $u \in \bar{r}$, ($n = 1, 2, \dots$), are non-degenerate surfaces, that $\lim_{n \rightarrow \infty} S_n = S$, that $x(u)$ and $x_n(u)$ are all non-degenerate, and that $x_n(u)$ approaches $x(u)$ uniformly. Suppose $x = X_n(U)$, $U \in \bar{R}$, is a representation of S_n satisfying: (i) it is of class L ; (ii) one of the induced (T , § 4, Theorem 3, and Def. 8) continuous monotone transformations, $u = u_n(U)$, of \bar{R} into \bar{r} carries three fixed (independently of n) distinct points, A , B , and C , of R^* into three fixed distinct points, a , b , and c , respectively, of r^* ; (iii) there is a constant M , independent of n , such that*

$$\iint_R [|\partial X_n / \partial U|^2 + |\partial X_n / \partial V|^2] dU dV < M, \quad (n = 1, 2, \dots).$$

Then the $X_n(U)$ are equicontinuous on \bar{R} .

If the $X_n(U)$ are not normalized on the boundary (i.e., do not satisfy (ii), they are equicontinuous on any closed set interior to R .

²⁴ A surface is said to be *non-degenerate* if it possesses a non-degenerate representation.

²⁵ See for instance, C. Caratheodory, "Conformal representation," *Cambridge Tracts in Mathematics and Mathematical Physics*, no. 28, § 161 and §§ 125-130 particularly.

²⁶ C. B. Morrey, Jr., "A class of representations of manifolds (Part II)," *American Journal of Mathematics*, vol. 56, no. 2 (1934), pp. 275-293.

In the normalized case, any limit function, $X(u)$, will satisfy all three conditions, and in the second case any limit function (defined over all of R) will satisfy (i) and (iii) on every closed region interior to R .

Proof. It is clear (Theorem 4, § 1) that we may take \bar{r} to be the unit circle, and a , b , and c to be equally spaced. On account of Lemma 3, we may also take \bar{R} to be the unit circle and A , B , and C to be equally spaced.²⁷ It is clearly sufficient to show that the functions $u_n(U)$ are equicontinuous.

We wish to observe at the outset that if P^*_1 and P^*_2 are points of R^* on a closed large (small) arc bounded by two fixed points and containing (not containing) the third, and we choose $\widehat{P^*_1 P^*_2}$ as that arc bounded by P^*_1 and P^*_2 and lying in the above large (small) arc, then all the points of the arc $\widehat{P^*_1 P^*_2}$ are carried into the corresponding arc $\widehat{p^*_1 p^*_2}$, $p^*_i = T_n(P^*_i)$, ($i = 1, 2$). This follows from the normalization of the T_n and the nature of the continua of R^* which are carried into points of r^* by a continuous monotone transformation.²⁸ Thus $|u_n(P^*_1) - u_n(P^*_2)|$ is equal to the oscillation of $u_n(U)$ on an arc $\widehat{P^*_1 P^*_2}$ which contains at most one of the fixed points A , B , C , unless this oscillation is equal to 2 in which case the above expression is not less than $3\frac{1}{2}$.

Let $C(P_0, \rho)$ denote the circle with any center at P_0 and radius ρ , $3\frac{1}{2}/2 = d > \rho > 0$. Suppose that the oscillation of some $u_n(U)$ in $\bar{C}(P_0, \rho_0) \cdot \bar{R} \geq \epsilon$, $2 \geq \epsilon > 0$, $d > \rho_0 > 0$. Then, from T, § 3, Theorem 1,²⁸ it is clear that the oscillation of $u_n(U)$ on $[C(P_0, \rho) \cdot R]^* \geq \epsilon$, $d > \rho \geq \rho_0$. Define $C(\rho)$ to be the arc of $[C(P_0, \rho)]^*$ which lies in \bar{R} , and let $P^*_{1\rho}$ and $P^*_{2\rho}$ be its end points, if they exist, in which case we let $\widehat{P^*_{1\rho} P^*_{2\rho}}$ be the arc $[C(P_0, \rho) \cdot R]^* \cdot R^*$. Then it is clear that the oscillation of $u_n(U)$ over $C(\rho) \geq \epsilon/2$, $d > \rho \geq \rho_0$, for (i) if the oscillation over $\widehat{P^*_{1\rho} P^*_{2\rho}}$ (which may be null or a point) $\leq \epsilon/2$, this is obviously the case, and (ii) if the oscillation over $\widehat{P^*_{1\rho} P^*_{2\rho}} > \epsilon/2$, then the oscillation over $C(\rho) \geq |u_n(P^*_{1\rho}) - u_n(P^*_{2\rho})|$ which is \geq the smaller of the numbers $3\frac{1}{2}$ and the oscillation of $u_n(U)$ over $\widehat{P^*_{1\rho} P^*_{2\rho}}$ (since $\widehat{P^*_{1\rho} P^*_{2\rho}}$ obviously cannot contain more than one of the fixed points), both of which exceed $\epsilon/2$, since $\epsilon \leq 2$.

Now, by Theorem 5, § 1, we can find an $\eta_1(\epsilon)$, positive with ϵ , such that,

²⁷ The argument can be carried through if R is any (Jordan) region, however.

²⁸ Or it follows directly from the theorem (T, § 3, Theorem 2) that a monotone transformation is the uniform limit (in the sense that the vector functions approach their limit uniformly) of a sequence of 1—1 continuous transformations, the statement being obvious for these.

for each $\epsilon > 0$, $0 < \eta_1(\epsilon) \leq \eta_1(\epsilon, x)$, $0 < \eta_1(\epsilon) \leq \eta_1(\epsilon, x_n)$, ($n = 1, 2, \dots$). Hence, if we define

$$\delta(\epsilon) = 2d \cdot k(\epsilon), \quad k(\epsilon) = e^{-(2\pi M/[\eta_1(\epsilon/2)]^2)},$$

we see that $|u_n(U_1) - u_n(U_2)| < \epsilon$, when $|U_1 - U_2| < \delta(\epsilon)$. For if this is not the case for some $u_n(U)$ and points U_1 and U_2 , the oscillation of $u_n(U)$ in $\bar{C}(P_0, kd) \cdot \bar{R} \geq \epsilon$, where $U_0 = (U_1 + U_2)/2$. Then for every ρ , $d > \rho \geq kd$, the oscillation of $u_n(U)$ on $C(\rho) \geq \epsilon/2$, and thus the oscillation of $X_n(U) = x_n[u_n(U)] \geq \eta_1(\epsilon/2)$ on $C(\rho)$. Let us choose polar coördinates with pole at P_0 (notice Lemma 2), and let $\theta_{1\rho}$ and $\theta_{2\rho}$, $2\pi \geq \theta_{2\rho} - \theta_{1\rho} > 0$ ($2\pi/3$ in fact), be the angular coördinates of $P_{1\rho}^*$ and $P_{2\rho}^*$ respectively (chosen so that $C(\rho)$ is the arc $\theta_{1\rho} \leq \theta \leq \theta_{2\rho}$) if they exist, otherwise let $\theta_{1\rho} = 0$, $\theta_{2\rho} = 2\pi$. Then using Schwartz's inequality

$$\begin{aligned} M &> \iint \left[\left| \frac{\partial X_n}{\partial U} \right|^2 + \left| \frac{\partial X_n}{\partial V} \right|^2 \right] dU dV \geq \int_{kd}^d \frac{d\rho}{\rho} \cdot \int_{\theta_{1\rho}}^{\theta_{2\rho}} \left| \frac{\partial X_n}{\partial \theta} \right|^2 d\theta \\ &\geq \int_{kd}^d \frac{d\rho}{\rho} \cdot \frac{1}{\theta_{2\rho} - \theta_{1\rho}} \left[\int_{\theta_{1\rho}}^{\theta_{2\rho}} \left| \frac{\partial X_n}{\partial \theta} \right| d\theta \right]^2 \geq \frac{[\eta_1(\epsilon/2)]^2}{2\pi} \log \frac{1}{k} = M \end{aligned}$$

which is impossible.

If we choose $r_0 < 1$ and $d = 1 - r_0$, the above argument demonstrates the equicontinuity of the $u_n(U)$ in the closed circle $U^2 + V^2 \leq r_0^2$ independently of the normalization on $U^2 + V^2 = 1$. This demonstrates the second statement in the conclusion of the theorem. The third statement follows immediately from Lemma 1.

THEOREM 2. *A necessary and sufficient condition that a non-degenerate surface, S , be of finite Lebesgue area is that it possess a generalized conformal (normalized) representation on the closed unit circle.*

Proof. The sufficiency of the condition is immediate from Lemmas 4 and 5.

To prove the existence of such a map, let $\{\bar{\Pi}_n\}$ be a sequence of polyhedra approaching S , where $\lim_{n \rightarrow \infty} L(\bar{\Pi}_n) = L(S)$, and $x = x(u)$ be a non-degenerate representation of S on \bar{r} . It is clear that we may replace each $\bar{\Pi}_n$ by a non-degenerate polyhedron, Π_n , such that $|L(\bar{\Pi}_n) - L(\Pi_n)| < 1/n$ and $\|\Pi_n, \bar{\Pi}_n\| < 1/n$ by merely moving the vertices²⁹ of $\bar{\Pi}_n$ slightly. Then, let

²⁹ By definition (given for instance in C. B. Morrey, Jr., *loc. cit.* R) $\bar{\Pi}_n$ can be represented on \bar{Q} (the unit square) by a function $\bar{x}_n(u)$ which is linear in triangles (a finite number of them). The vertices of $\bar{\Pi}_n$ are merely the points corresponding to the vertices of the triangles in \bar{Q} .

$x = x_n(u)$ be a sequence of non-degenerate representations of Π_n such that $x_n(u)$ approaches $x(u)$ uniformly (that this is possible follows from T, § 5, Theorem 2). Then, let a, b , and c be three distinct points of r^* , and A, B , and C three distinct points of R^* , where \bar{R} is the closed unit circle. Let $x = X_n(U)$, $U \in \bar{R}$, be a generalized conformal representation of Π_n on \bar{R} so that an induced transformation, $u = u_n(U)$ of \bar{R} into \bar{r} , carries A, B , and C into a, b , and c respectively. By Lemma 5, and the conformality,

$$L(\Pi_n) = (1/2) \int \int_r [|\partial X_n / \partial U|^2 + |\partial X_n / \partial V|^2] dU dV.$$

Thus, the hypotheses of Theorem 1 are fulfilled and thus we may extract a subsequence of the $X_n(U)$ which converges uniformly to a function $X(U)$. Clearly $x = X(U)$ is a representation of S^{30} and, by Lemma 7, it is generalized conformal.

The following very important theorem due to McShane³¹ and used by him in his very interesting solution of the problem of Plateau is a consequence of the above theorem and the theorem of T, § 5, Theorem 5.

THEOREM 3. *Let S be a Lebesgue monotone (T, § 5, Def. 4) surface of finite area bounded by a Jordan curve. Then S possesses a generalized conformal representation on the unit circle in which three given distinct points on the boundary of S correspond to three given distinct points on the circumference of the unit circle.*

THE UNIVERSITY OF CALIFORNIA.

³⁰ For let S_1 be this surface. Clearly $\lim_{n \rightarrow \infty} \|S, S_n\| = \lim_{n \rightarrow \infty} \|S_1, S_n\| = 0$. But since the Fréchet distance satisfies the "triangle inequality," $\|S, S_1\| \leq \|S, S_n\| + \|S_1, S_n\|$, we see that $\|S, S_1\| = 0$ and thus $S = S_1$.

³¹ E. J. McShane, *loc. cit.* (1st ref.).

5000000
* LIBRARY

POSTULATES FOR BOOLEAN ALGEBRAS AND GENERALIZED BOOLEAN ALGEBRAS.

By M. H. STONE.

In this paper we present a new set of postulates for Boolean algebras in terms of logical addition and logical multiplication (or, to use the language of the algebra of classes, in terms of union and intersection), a similar set of postulates for generalized Boolean algebras (distinguished by the possible absence of a unit or "universe" element), and an investigation of the independence of the proposed postulates. This material was originally designed to constitute an introduction to a general theory of Boolean algebras, to be developed by the methods appropriate to modern concepts of algebra.¹ The postulates were therefore chosen with the intention of emphasizing the analogy between Boolean algebras and abstract rings, the latter being systems which have already undergone extensive analysis. From this point of view our postulates appear to be as satisfactory as possible, so long as logical addition and multiplication are to be treated as the analogues of ring addition and multiplication. We have recently observed, however, that Boolean algebras can be regarded as rings of special type when the operation of forming the symmetric difference is taken as ring addition.² In consequence, the most natural approach to a mathematical theory of Boolean algebras is not one based upon the material of the present paper. Since the results which are collected here seem to possess some intrinsic interest—in particular, they offer a simplification of Huntington's set of postulates in terms of logical addition and multiplication, and also establish the redundancy of one of the postulates in the set proposed by De Morgan³—we venture to present them apart from our general theory of Boolean algebras.

1. *A new set of postulates for Boolean algebras.* We shall define a Boolean algebra as a system with double composition in which the operations

¹ For a brief sketch of the theory, see *Bulletin of the American Mathematical Society*, abstract No. 39-3-86; *Proceedings of the National Academy of Sciences*, vol. 20 (1934), pp. 197-202.

² See *Proceedings of the National Academy of Sciences*, vol. 21 (1935), pp. 103-105; the formal definition of the symmetric difference in terms of logical addition and multiplication is given as Definition 4 below.

³ See § 2 below.

of addition and multiplication are required to satisfy certain further postulates. In selecting these postulates we are guided by three principles: in the first place, they shall embody known properties of the operations of forming finite unions and intersections of classes; in the second place, they shall embody so far as possible only such properties as are valid of the operations of addition and multiplication in a ring; and, in the third place, they shall be independent. We recall ⁴ that a ring is a system with double composition in which addition is commutative, both addition and multiplication are associative, multiplication is both left-distributive and right-distributive with respect to addition, and the equation ⁵ $x \vee a = b$ has a solution for arbitrary elements a and b . It is easily shown that in an arbitrary ring this equation has a unique solution, and that there exists a unique element 0 such that $a \vee 0 = 0 \vee a = a$ for every element a . In an algebra of classes, the operations of forming the finite union and finite intersection have all these properties except that the equation $x \vee a = b$ does not have a solution unless $a \subset b$ and does not have a unique solution, even when this condition is satisfied, unless the auxiliary relation $xa = 0$ be required. The algebra of classes is further differentiated by the existence of the special properties expressed by the equations $a \vee a = a$ and $aa = a$. The second of these equations is not generally true for rings but can be verified without exception in certain special rings. On the other hand, the only ring element which has the property that $a \vee a = a$ is the element 0 described above. In accordance with the first two principles stated above, we may postulate the commutative law for addition, the associative law for addition and multiplication, the two distributive laws for multiplication with respect to addition, the existence of an element 0 such that $a \vee 0 = a$ for every element a , the existence of a solution of the simultaneous equations $x \vee a = e$ and $xa = 0$ for every element a and a certain fixed element e , and the two special rules $a \vee a = a$ and $aa = a$. Every algebra of classes has all these properties; there are special rings which have all these properties except the one expressed by the equation $a \vee a = a$; and the only ring which has all these properties is the trivial ring consisting of the single element 0 . Our third principle, however, requires us to reject the associative laws for addition and multiplication, since they can be deduced from the other properties listed. Thus our definition of Boolean algebras can be stated as follows:

DEFINITION 1. *A Boolean algebra A is a system with double composition which satisfies the further postulates:*

⁴ See B. L. van der Waerden, *Moderne Algebra*, Bd. 1 (Berlin, 1930), pp. 36-39.

⁵ We shall denote addition by the symbol \vee .

Postulate 1₁. $a \vee b = b \vee a$;

Postulate 3₁. $a(b \vee c) = ab \vee ac$;

Postulate 3₂. $(a \vee b)c = ac \vee bc$;

Postulate 4₁. *There exists an element 0 in A such that $a \vee 0 = a$ for every element a in A;*

* Postulate 5. *If there exists an element 0 with the property required by Postulate 4₁, there exists at least one such element 0 to which corresponds a fixed element e in A such that the simultaneous equations $x \vee a = e$, $ax = 0$ have a solution for every element a in A;*

Postulate 6₁. $a \vee a = a$;

Postulate 6₂. $aa = a$.

Any element such as that postulated in Postulate 4₁ is called a zero; and any element such as the element e postulated in Postulate 5 is called a unit.

The numbering of the postulates has been chosen with a view to certain subsequent modifications in the system of postulates. We may point out that, in accordance with Postulate 4₁, a Boolean algebra contains at least one element. It is often required that a Boolean algebra have at least two elements; but for our purposes it is preferable to admit as a Boolean algebra a one-element system, consisting of zero elements alone. We call attention to the fact that the replacement of the simultaneous equations of Postulate 5 by the simultaneous equations $x \vee a = e$, $ax = 0$ does not constitute an essential alteration: if we rewrite the altered set of postulates in terms of a new multiplication \circ defined by writing $a \circ b$ for ba we obtain the original set once more.

We must now proceed to deduce from the postulated properties of the operations in a Boolean algebra a series of further properties, including, for instance, the associative law for addition and multiplication and the commutative law for multiplication. In order to condense the demonstrations, we shall arrange them so far as possible in the form of continued equations, citing the propositions used in passing across successive equality signs by successive groups set off by commas. In the citations, we shall replace the words postulate, definition, theorem, lemma, and hypothesis by their initial letters P, D, T, L, and H respectively. Those propositions which do not hold in generalized Boolean algebras, introduced later, will be indicated by prefixing an asterisk.

THEOREM 1. *The element 0 of Postulate 4₁ is unique.*

If 0_1 and 0_2 are elements with the property postulated in P 4₁, then $0_1 = 0_1 \vee 0_2 = 0_2 \vee 0_1 = 0_2$ by P 4₁, P 1₁, P 4₁.

* THEOREM 2. *If 0 and e are elements with the properties demanded in * Postulate 5, then $ea = a$ for every element a.*

Let x be an element such that $x \vee a = e$, $xa = 0$ in accordance with * P 5. Then by *P 5, P 3₂, *P 5 — P 6₂, P 1₁, P 4₁ we have $ea = (x \vee a)a = xa \vee aa = 0 \vee a = a \vee 0 = a$.

* LEMMA 1. *If a' is any solution of the equations $x \vee a = e$, $xa = 0$, then $a'e = a'$.*

By * P 5, P 3₁, P 6₂ — * P 5, P 4₁, we have $a'e = a'(a' \vee a) = a'a' \vee a'a = a' \vee 0 = a'$.

* LEMMA 2. *If a' is any solution of the equations $x \vee a = e$, $xa = 0$, and if a'' is any solution of the equations $x \vee a' = e$, $xa' = 0$, then $a'' = a$.*

By *L 1, *P 5, P 3₁, *P 5, *P 5, P 3₂, P 1₁, *P 5, *T 2, we have

$$\begin{aligned} a'' &= a'e = a''(a' \vee a) = a''a' \vee a''a = 0 \vee a''a \\ &= a'a \vee a''a = (a' \vee a'')a = (a'' \vee a')a = ea = a. \end{aligned}$$

* LEMMA 3. *If a' is any solution of the equations $x \vee a = e$, $xa = 0$, then $aa' = 0$.*

By *L 2, we have $aa' = a''a' = 0$.

* THEOREM 3. *If 0 and e are elements with the properties demanded in * Postulate 5, then $ae = a$ for every element a.*

Let a' be any solution of the equations $x \vee a = e$, $xa = 0$. Then by *P 5, P 3₁, *L 3 — P 6₂, P 1₁, P 4₁ we have $ae = a(a' \vee a) = aa' \vee aa = 0 \vee a = a \vee 0 = a$.

* THEOREM 4. *The element e of Postulate 5 is unique.*

If e_1 and e_2 are any two elements with the property postulated in *P 5, whether they correspond to the same or equal zero elements, we have $e_1 = e_1e_2 = e_2$ by *T 3, *T 2.

* THEOREM 5. *If e is the element described in * Postulate 5, then $e \vee a = a \vee e = e$ for every element a.*

Let a' be any solution of the equations $x \vee a = e$, $xa = 0$. Then, by P 1₁, *T 2, *P 5, P 3₂, P 3₁, *P 5 — *T 4 — P 6₂ — *T 4, P 1₁ — P 4₁ — P 6₁, *P 5, we have

$$\begin{aligned} e \vee a &= a \vee e = e(a \vee e) = (a' \vee a)(a \vee e) = a'(a \vee e) \vee a(a \vee e) \\ &= (a'a \vee a'e) \vee (aa \vee ae) = (0 \vee a') \vee (a \vee a) = a' \vee a = e. \end{aligned}$$

THEOREM 6. If 0 is the element of Postulate 4, then $a0 = 0a = 0$ for every element a .

By P 4₁, *T 2, P 3₂, *T 5, *T 2, we have $a0 = a0 \vee 0 = a0 \vee e0 = (a \vee e)0 = e0 = 0$; and, by P 4₁, *T 3, P 3₁, *T 5, *T 3, we have $0a = 0a \vee 0 = 0a \vee 0e = 0(a \vee e) = 0e = 0$.

*THEOREM 7. The equations $x \vee a = e$, $xa = 0$ of *Postulate 5 have a unique solution.

Let $0_1, 0_2$ be zero elements; let e_1 and e_2 be corresponding unit elements; let a'_1 be any solution of the equations $x \vee a_1 = e_1$, $xa_1 = 0_1$; let a'_2 be any solution of the equations $x \vee a_2 = e_2$, $xa_2 = 0_2$; and let $a_1 = a_2$. By *T 3, *P 5, H, P 3₁, *P 5, T 1, *L 3, P 3₂, H, *P 5, *T 2, we then have

$$\begin{aligned} a'_1 &= a'_1 e_2 = a'_1 (a'_2 \vee a_2) = a'_1 (a'_2 \vee a_1) = a'_1 a'_2 \vee a'_1 a_1 = a'_1 a'_2 \vee 0_1 \\ &= a'_1 a'_2 \vee 0_2 = a'_1 a'_2 \vee a_2 a'_2 = (a'_1 \vee a_2) a'_2 = (a'_1 \vee a_1) a'_2 = e_1 a'_2 = a'_2. \end{aligned}$$

In view of *Theorem 7, *Lemma 2, it is convenient to introduce the following definition and theorem.

*DEFINITION 2. The unary operation which takes the element a into the solutions of the equations of *Postulate 5, namely, the simultaneous equations $x \vee a = e$ and $xa = 0$, is called complementation; and the element which results from this operation is denoted by a' and is called the complement of a . The element $(a')'$ is denoted by a'' .

*THEOREM 8. The operation of complementation has the following properties:

$$(1) a = b \text{ implies } a' = b'; \quad (2) a'' = a; \quad (3) 0' = e; \quad (4) e' = 0.$$

This theorem merely summarizes the results of *Theorem 7 and *Lemma 2 and the special cases $e \vee 0 = e$, $e0 = 0$ and $0 \vee e = e$, $0e = 0$ of the relations proved in *Theorems 5 and 6.

*LEMMA 4₁. If $a \vee b = 0$, then $a'b = 0$.

By P 4₁, *P 5, P 3₁ — P 1₁, H, T 6, we have

$$a'b = a'b \vee 0 = a'b \vee a'a = a'(a \vee b) = a'0 = 0.$$

* LEMMA 4₂. If $a \vee b = 0$, then $ab' = 0$.

By P 4₁, *L 3, P 3₂, H, T 6, we have

$$ab' = ab' \vee 0 = ab' \vee bb' = (a \vee b)b' = 0b'$$

THEOREM 9. If $a \vee b = 0$, then $a = b = 0$.

By P 1₁ and T 1, it is sufficient to show that $a =$
*P 5, P 3₁, *L 4₂, *L 4₁, P 3₂, *P 5, *T 2, H, we have

$$\begin{aligned} a &= a \vee a = a \vee ae = a \vee a(b' \vee b) = a \vee (ab' \vee ab) \\ &= a \vee (0 \vee ab) = a \vee (a'b \vee ab) = a \vee (a' \vee a)b = \end{aligned}$$

* LEMMA 5₁. If $c = (a \vee b)'$, then $ca = 0$.

By P 3₁, *P 5, we have $ca \vee cb = c(a \vee b) = 0$. By
 $ca = 0$,

* LEMMA 5₂. If $c = (a \vee b)'$, then $ac = 0$.

By P 3₂, *L 3, we have $ac \vee bc = (a \vee b)c = 0$. By
 $ac = 0$.

* LEMMA 5₃. If $c = (a \vee b)'$, then $ca' = c$.

By P 4₁, *L 5₁, P 3₁, *P 5, *T 3, we have

$$ca' = ca' \vee 0 = ca' \vee ca = c(a' \vee a) = ce =$$

* LEMMA 5₄. If $c = (a \vee b)'$, then $a'c = c$.

By P 4₁, *L 5₂, P 3₂, *P 5, *T 2, we have

$$a'c = a'c \vee 0 = a'c \vee ac = (a' \vee a)c = ec$$

* LEMMA 5₅. If $c = (a \vee b)'$, then $a' = c \vee ba'$.

By *T 3, *P 5, P 3₁, *L 5₄ — P 3₁, *P 5, P 1₁, P 4₁,
 $a' = a'e = a'[c \vee (a \vee b)] = a'c \vee a'(a \vee b)$
 $= c \vee (a'a \vee a'b) = c \vee (0 \vee a'b) = c$

* LEMMA 5₆. If $c = (a \vee b)'$, then $a' = c \vee ba'$.

By *T 2, *P 5, P 3₂, *L 5₃ — P 3₂, *L 3, P 1₁, P 4₁
 $a' = ea' = [c \vee (a \vee b)]a' = ca' \vee (a \vee b)a'$
 $= c \vee (aa' \vee ba') = c \vee (0 \vee ba') = c$

* LEMMA 6₁. $a(a'b) = 0$.

Let $c = (a \vee b)'$. Then, by P 4₁, *L 5₂, P 3₁, P 1₁, *L 5₅, *L 3, we have
 $a(a'b) = a(a'b) \vee 0 = a(a'b) \vee ac = a(a'b \vee c) = a(c \vee a'b) = aa' = 0$.

* LEMMA 6₂. $a(ba') = 0$.

Let $c = (a \vee b)'$. Then, by P 4₁, *L 5₂, P 3₁, P 1₁, *L 5₅, *L 3, we have
 $a(ba') = a(ba') \vee 0 = a(ba') \vee ac = a(ba' \vee c) = a(c \vee ba') = aa' = 0$.

* LEMMA 6₃. $(a'b)a = 0$.

Let $c = (a \vee b)'$. Then, by P 4₁, *L 5₁, P 3₂, P 1₁, *L 5₅, *P 5, we have
 $(a'b)a = (a'b)a \vee 0 = (a'b)a \vee ca = (a'b \vee c)a = (c \vee a'b)a = a'a = 0$.

* LEMMA 6₄. $(ba')a = 0$.

Let $c = (a \vee b)'$. Then, by P 4₁, *L 5₁, P 3₂, P 1₁, *L 5₅, *P 5, we have
 $(ba')a = (ba')a \vee 0 = (ba')a \vee ca = (ba' \vee c)a = (c \vee ba')a = a'a = 0$.

THEOREM 10. $a \vee ab = a \vee ba = a$.

Let $c = (a \vee b)'$. Then, by P 6₂ — P 4₁, P 3₁ — *L 5₂, P 3₁, P 1₁, *P 5, *T 3, *T 2, *P 5, P 3₂, *L 5₁ — P 3₂, P 6₂ — P 1₁, P 4₁, we have

$$\begin{aligned} a \vee ab &= (aa \vee ab) \vee 0 = a(a \vee b) \vee ac = a[(a \vee b) \vee c] = a[c \vee (a \vee b)] = ae \\ &= a = ea = [c \vee (a \vee b)]a = ca \vee (a \vee b)a = 0 \vee (aa \vee ba) = (a \vee ba) \vee 0 = a \vee ba. \end{aligned}$$

* LEMMA 7. $a \vee a'b = a \vee b$.

By T 10 — P 4₁, P 1₁, P 6₂ — *P 5, P 3₁, P 3₁, P 1₁, *P 5, *T 2, we have
 $a \vee a'b = (a \vee ab) \vee (a'b \vee 0) = (a \vee ab) \vee (0 \vee a'b) = (aa \vee ab) \vee (a'a \vee a'b)$
 $= a(a \vee b) \vee a'(a \vee b) = (a \vee a')(a \vee b) = (a' \vee a)(a \vee b) = e(a \vee b) = a \vee b$.

LEMMA 8₁. $(a \vee b)(a \vee c) = a \vee (ba \vee bc)$.

By P 3₂, P 3₁, P 6₂, T 10, we have

$$\begin{aligned} (a \vee b)(a \vee c) &= a(a \vee c) \vee b(a \vee c) \\ &= (aa \vee ac) \vee (ba \vee bc) = (a \vee ac) \vee (ba \vee bc) = a \vee (ba \vee bc). \end{aligned}$$

LEMMA 8₂. $a[a \vee (ba \vee bc)] = a.$

By P 3₁, P 6₂, T 10, we have

$$a[a \vee (ba \vee bc)] = aa \vee a(ba \vee bc) = a \vee a(ba \vee$$

* LEMMA 8₃. $a'[a \vee (ba \vee bc)] = a'(bc).$

By P 3₁, *P 5, P 1₁, P 4₁ — P 3₁, *L 2, *L 6₂, I

$$\begin{aligned} a'[a \vee (ba \vee bc)] &= a'a \vee a'(ba \vee bc) = 0 \vee a'(ba \vee bc) = a'(b \\ &= a'(ba) \vee a'(bc) = a'(ba'') \vee a'(bc) = 0 \vee a'(bc) = a'(b \end{aligned}$$

THEOREM 11. $(a \vee b)(a \vee c) = a \vee bc.$

By L 8₁, *T 2, *P 5, P 3₂, *L 8₃ — L 8₂, P 1₁, *L 7, W

$$\begin{aligned} (a \vee b)(a \vee c) &= a \vee (ba \vee bc) = e[a \vee (ba \vee bc)] = (a' \vee a)[\\ &= a'[a \vee (ba \vee bc)] \vee a[a \vee (ba \vee bc)] = a'(bc) \vee a = a \vee a \end{aligned}$$

* LEMMA 9. $(a \vee b) \vee a' = e.$

By *T 2, *P 5, P 3₂, P 3₁, P 3₁ — P 6₂ — P 3₁ — *L 3, P 1₁ — T 10, P 4₁, P 1₁, T 10, *P 5, we have

$$\begin{aligned} (a \vee b) \vee a' &= e[(a \vee b) \vee a'] = (a' \vee a)[(a \vee b) \vee a'] \\ &= a'[(a \vee b) \vee a'] \vee a[(a \vee b) \vee a'] = [a'(a \vee b) \vee a'a'] \vee [a \\ &= [(a'a \vee a'b) \vee a'] \vee [(aa \vee ab) \vee 0] = [(0 \vee a'b) \vee a'] \vee (\\ &= [(a'b \vee 0) \vee a'] \vee a = (a'b \vee a') \vee a = (a' \vee a'b) \vee a = a' \end{aligned}$$

* THEOREM 12₁. $(a \vee b)' = a'b'.$

By *T 7, it is sufficient to prove that $a'b'$ is a solution equations $x \vee (a \vee b) = e$, $x(a \vee b) = 0$. By P 1₁, T 11, P 6₂, we have

$$a'b' \vee (a \vee b) = (a \vee b) \vee a'b' = [(a \vee b) \vee a'][(a \vee b) \vee b'] = e[$$

and, by P 3₁, *L 6₃ — *L 6₄, P 6₁, we have also

$$a'b'(a \vee b) = (a'b')a \vee (a'b')b = 0 \vee 0 =$$

* THEOREM 12₂. $ab = (a' \vee b')'$ and $(ab)' = a' \vee b'.$

By *L 2, *T 12₁, we have $ab = a''b'' = (a' \vee b')$
 $(ab)' = (a' \vee b')'' = a' \vee b'.$

THEOREM 13. $ba = ab$.

By *T 12₂, P 1₁, *T 12₂, we have $ba = (b' \vee a')' = (a' \vee b')' = ab$.

LEMMA 10₁. $a[(a \vee b) \vee c] = a[a \vee (b \vee c)] = a$.

By P 3₁, P 3₁, P 6₂, T 10, T 10, T 10, P 6₂, P 3₁, we have

$$\begin{aligned} a[(a \vee b) \vee c] &= a(a \vee b) \vee ac = (aa \vee ab) \vee ac = (a \vee ab) \vee ac \\ &= a \vee ac = a = a \vee a(b \vee c) = aa \vee a(b \vee c) = a[a \vee (b \vee c)]. \end{aligned}$$

*LEMMA 10₂. $a'[(a \vee b) \vee a] = a'[a \vee (b \vee c)] = a'(b \vee c)$.

By P 3₁, P 3₁, *P 5, P 1₁, P 4₁, P 3₁, P 4₁, P 1₁, *P 5, P 3₁, we have

$$\begin{aligned} a'[(a \vee b) \vee c] &= a'(a \vee b) \vee a'c = (a'a \vee a'b) \vee a'c = (0 \vee a'b) \vee a'c \\ &= (a'b \vee 0) \vee a'c = a'b \vee a'c = a'(b \vee c) = a'(b \vee c) \vee 0 \\ &= 0 \vee a'(b \vee c) = a'a \vee a'(b \vee c) = a'[a \vee (b \vee c)]. \end{aligned}$$

THEOREM 14. $(a \vee b) \vee c = a \vee (b \vee c)$.

By *T 2, *P 5, P 3₂, L 10₁—*L 10₂, P 3₂, *P 5, *T 2, we have

$$\begin{aligned} (a \vee b) \vee c &= e[(a \vee b) \vee c] = (a' \vee a)[(a \vee b) \vee c] \\ &= a'[(a \vee b) \vee c] \vee a[(a \vee b) \vee c] = a'[a \vee (b \vee c)] \vee a[a \vee (b \vee c)] \\ &= (a' \vee a)[a \vee (b \vee c)] = e[a \vee (b \vee c)] = a \vee (b \vee c). \end{aligned}$$

THEOREM 15. $(ab)c = a(bc)$.

By *T 12₂, *T 12₂, T 14, *T 12₂, *T 12₂, we have

$$(ab)c = ((ab)' \vee c')' = ((a' \vee b') \vee c')' = (a' \vee (b' \vee c'))' = (a' \vee (bc)')' = a(bc).$$

THEOREM 16₁. If the simultaneous equations $x \vee a = b$, $xa = 0$ have a solution c , then $ba = a$, $bc = c$, and the solution is unique.

If c is a solution, we have $ba = (c \vee a)a = ca \vee aa = 0 \vee a = a \vee 0 = a$ by H, P 3₂, H—P 6₂, P 1₁, P 4₁; and also $bc = (c \vee a)c = cc \vee ca = c \vee 0 = c$ by H, P 3₂—T 13, P 6₂—H, P 4₁. If c_1 and c_2 are solutions of the respective systems $x \vee a_1 = b_1$, $xa_1 = 0_1$ and $x \vee a_2 = b_2$, $xa_2 = 0_2$, where $a_1 = a_2$, $b_1 = b_2$, and $0_1 = 0_2$, we have to show that $c_1 = c_2$. For this it is sufficient, by what we have just proved, to show that $b_1c_1 = b_2c_2$. Now we have

$$\begin{aligned} b_1c_1 &= b_2c_1 = (c_2 \vee a_2)c_1 = (c_2 \vee a_1)c_1 = c_2c_1 \vee a_1c_1 = c_1c_2 \vee c_1a_1 \\ &= c_1c_2 \vee 0_1 = c_1c_2 \vee 0_2 = c_1c_2 \vee c_2a_2 = c_1c_2 \vee a_1c_2 = (c_1 \vee a_1)c_2 = b_1c_2 = b_2c_2 \end{aligned}$$

by H, H, H, P 3₂, T 13—T 13, H, H, H, T 13—H, P 3₂, H, H.

* THEOREM 16₂. *If $ba = a$, the equations of Theorem 16₁ have the solution ba' .*

We have $ba' \vee a = ba' \vee ba = b(a' \vee a) = be = b$ by H, P 3₁, *P 5, *T 3; and $(ba')a = 0$ by *L 6₄.

We have now obtained the principal rules governing the operations of addition and multiplication in a Boolean algebra. Both operations are commutative, by Postulate 1₁ and Theorem 13; both are associative, by Theorems 14 and 15; and each is distributive with respect to the other, by Postulates 3₁ and 3₂ and Theorem 11. The further rules $a \vee a = a$, $aa = a$, $a \vee ab = a$ are given by Postulate 6₁, Postulate 6₂, and Theorem 10 respectively. By virtue of the commutative and associative laws, we can simplify our notation: it is no longer necessary to indicate by the use of parentheses or by the ordering of the base elements the precise structure of an additive or multiplicative construct or polynomial in a Boolean algebra. In Definition 2 we have introduced the new unary operation of complementation; the principal rules which govern it are embodied in the relations $a'' = a$, $a' \vee a = e$, $a'a = 0$, $(a \vee b)' = a'b'$, and $(ab)' = a' \vee b'$ of Theorem 8, Definition 2, Definition 2, Theorem 12₁, and Theorem 12₂, respectively. The last two relations reveal clearly the complete duality between the operations of addition and multiplication, which one would suspect from a comparison of the rules they obey: it is evident that any rule connecting addition and multiplication can be converted, by taking complements, into a rule in which the rôles of the two operations are interchanged. A systematic treatment of this duality, while unnecessary for our purposes, can easily be supplied. Finally, we have presented in Postulate 5 and Theorems 16₁ and 16₂ those facts from the theory of equations which will be most useful in the sequel. Since any algebra of classes satisfies the postulates of Definition 1 or is a subsystem of an algebra with this property, we can apply these results to the operations upon classes; it is easily seen that the operation of complementation here becomes the familiar operation of forming the complement of a given class.

We shall now study the definition of a relation in a Boolean algebra which is analogous to the relation \subset defined for classes. We can introduce such a relation in any one of a number of equivalent ways. Here we choose to make our fundamental definition in terms of the operation of multiplication.

DEFINITION 3. *If $ab = a$, we write $a' < b$, $b > a$; and we say that a is less than b , b greater than a .*

THEOREM 17₁. $0 < a$ for every element a .

We have $0a = 0$ for every a by T 6.

*THEOREM 17₂. $a < e$ for every element a .

We have $ae = a$ for every a by *T 3.

THEOREM 18₁. If $a < b$ and $b < c$, then $a < c$.

We have $ac = (ab)c = a(bc) = ab = a$ by H, T 15, H, H.

THEOREM 18₂. If $a < b$, $b < c$, and $a = c$, then $a = b = c$.

We have $c = a = ab = cb = bc = b$ by H, H, H, T 13, H.

THEOREM 18₃. If $a < b$ and $b < a$, then $a = b$; and conversely.

If $a < b$ and $b < a$, we have $a = a$ and hence $a = b$ by T 18₂. If $a = b$, we have $ab = aa = a$ by P 6₂; and $ba = bb = b$ by P 6₂.

THEOREM 19. If $a < b$ and $c < d$, then $ac < bd$ and $a \vee c < b \vee d$.

We have $(ac)(bd) = (ab)(cd) = ac$ by T 13 — T 15, H; and $(a \vee c)(b \vee d) = (ab \vee ad) \vee (cb \vee cd) = (a \vee ad) \vee (cb \vee c) = a \vee c$ by P 3₁ — P 3₂, H, T 10 — P 1₁ — T 10.

THEOREM 20₁. In order that $a < b$ and $a < c$, it is necessary and sufficient that $a < bc$.

In order to prove the sufficiency of the condition $a < bc$, it is sufficient by T 18₁ to show that $bc < b$ and $bc < c$. Now we have $(bc)b = b(bc) = (bb)c = bc$ by T 13, T 15, P 6₂; and $(bc)c = b(cc) = bc$ by T 15, P 6₂. In order to prove the necessity of the condition $a < bc$, we have merely to note that $a(bc) = (aa)(bc) = (ab)(ac) = aa = a$ by P 6₂, T 13 — T 15, H — H, P 6₂.

THEOREM 20₂. In order that $a > b$ and $a > c$, it is necessary and sufficient that $a > b \vee c$.

In order to prove the sufficiency of the condition $a > b \vee c$, it is sufficient by T 18₁ to show that $b \vee c > b$ and $b \vee c > c$. Now we have $b > b$ by T 18₃, $c > 0$ by T 17₁, $b \vee c > b \vee 0$ by T 19, $b \vee 0 > b$ by P 4₁ — T 18₃, and $b \vee c > b$ by T 18₁; and $b \vee c > c \vee b$ by P 1₁ — T 18₃, $c \vee b > c$ by what has just been proved, and $b \vee c > c$ by T 18₁. In order to prove the necessity of the condition $a > b \vee c$, we note that $a \vee a > b \vee c$ by H — H — T 19, $a > a \vee a$ by P 6₁ — T 18₃, and $a > b \vee c$ by T 18₁.

THEOREM 21. *In order that $a < b$, it is necessary and sufficient that $a \vee b = b$.*

If $ab = a$, we have $a \vee b = ab \vee b = b \vee ab = b$ by H, P 1₁, T 10; and, if $a \vee b = b$, we have $ab = a(a \vee b) = aa \vee ab = a \vee ab = a$ by H, P 3₁, P 6₂, T 10.

*THEOREM 22. *In order that $a < b$, it is necessary and sufficient that $a' > b'$.*

If $a < b$, we have $b'a' = a'b' = (a \vee b)' = b'$ by T 13, *T 12₁, T 21; and, if $a' > b'$, we have $ab = ba = (b' \vee a')' = (a')' = a'' = a$ by T 13, *T 12₂, T 21, *D 2, *T 8.

*THEOREM 23₁. *In order that $a < b$, it is necessary and sufficient that $ab' = 0$.*

If $ab = a$, we have $ab' = (ab)b' = 0$ by H, *L 2 — *L 6₄; and if $ab' = 0$, we have $ab = ab \vee 0 = ab \vee ab' = a(b \vee b') = ae = a$ by P 4₁, H, P 3₁, P 1₁ — *P 5, *T 3.

*THEOREM 23₂. *In order that $a < b$, it is necessary and sufficient that $a' \vee b = e$.*

If $a < b$, then $a' \vee b = a' \vee b'' = (ab')' = 0' = e$ by *T 8, *T 12₂, *T 23₁, *T 8; and, if $a' \vee b = e$, then $ab' = (a' \vee b'')' = (a' \vee b)' = e' = 0$ by *T 12₂, *T 8, H, *T 8, so that $a < b$ by *T 23₁.

We have now examined in sufficient detail the relation introduced in Definition 3. We have not distinguished the case where we have $a < b$ and $a \neq b$ from the case where we have merely $a < b$; in the present paper it is not important that we do so. We have not proved, nor can we do so in general, that one of the two relations $a < b$, $b < a$ holds whatever the elements a and b . Thus, in view of Theorem 18₁, we may describe the relation indicated by the symbol $<$ as an incomplete ordering relation. The most striking and important connections between this relation and the operations of addition, multiplication, and complementation are those presented in Theorems 19 and 22.

We next study the introduction⁶ of an operation analogous to the operation of forming the symmetric difference of two classes—that is, the class of elements belonging to either one but not both of the given classes. In order to abbreviate the somewhat tedious proofs of several of the theorems which we have to establish—in particular, those of Theorems 32, 42, and 43—we shall

⁶ See Daniell, *Bulletin of the American Mathematical Society*, vol. 23 (1916), pp. 446-458.

occasionally be less precise than hitherto in citing earlier propositions. We commence with a variation upon Theorem 16₂.

THEOREM 24. *The equations $x \vee ab = a \vee b$, $x(ab) = 0$ have a unique solution.*

If these equations have a solution, it is unique, by Theorem 16₁. We have $(a \vee b)(ab) = (aa)b \vee a(bb) = ab \vee ab = ab$ by P 3₂ — T 13 — T 15, P 6₂, P 6₁. Hence we see that the equations in question have a solution, by *T 16₂.

DEFINITION 4. *The unique solution of the simultaneous equations $x \vee ab = a \vee b$, $x(ab) = 0$ is called the symmetric difference of the elements a and b , and is denoted by $a \Delta b$. The operation of forming the symmetric difference is called differentiation.⁷*

THEOREM 25. *If $a = c$ and $b = d$, then $a \Delta b = c \Delta d$.*

By hypothesis, we have $ab = cd$ and $a \vee b = c \vee d$. From T 24 and D 4, we conclude that $a \Delta b = c \Delta d$.

THEOREM 26. $b \Delta a = a \Delta b$.

We have $a \vee b = b \vee a$ by P 1₁ and $ab = ba$ by T 13. Hence we have $b \Delta a = a \Delta b$ by T 16₁ and D 4.

THEOREM 27. $a \Delta b = (a \vee b) \Delta (ab)$.

We have $(a \vee b) \vee ab = (a \vee ab) \vee b = a \vee b$ by P 1₁ — T 14, T 10; and $(a \vee b)(ab) = ab$ by the proof of T 24. Hence we conclude that $a \Delta b = (a \vee b) \Delta (ab)$ by T 16₁ and D 4.

THEOREM 28. *In order that $a \Delta b = 0$, it is necessary and sufficient that $a = b$; in particular, $a \Delta a = 0$ for every element a .*

If $a \Delta b = 0$, we have $a \vee b = (a \Delta b) \vee ab = 0 \vee ab = ab$ by D 4, H, P 1₁ — P 4₁; we have $ab < a$, $a < a \vee b$ by T 20₁, T 20₂ and $ab < b$, $b < a \vee b$ by T 20₁, T 20₂; hence we conclude that $a = a \vee b = b$ by T 18₂. If $a = b$, we have $a \vee b = a \vee a = a$ by P 6₁ and $ab = aa = a$ by P 6₂; hence $a \Delta b$ is the unique solution of the equations $x \vee a = a$, $xa = 0$ in accordance with T 24 and D 4; since these equations have the solution 0 by P 1₁, P 4₁, T 6, we conclude that $a \Delta b = 0$. Since $a = a$, we have in particular $a \Delta a = 0$.

⁷ This term is introduced on the basis of its colloquial logical connotations. There is little danger that it will be confused with the term "differentiation" used in mathematical analysis with a special technical meaning.

THEOREM 29. $a \Delta 0 = a$.

We have $a \vee 0 = a$ by P 4₁ and $a0 = 0$ by T 6; hence $a \Delta 0$ is the unique solution of the equations $x \vee 0 = a$, $x0 = 0$ in accordance with T 24 and D 4; since these equations have the solution a by P 4₁, T 6, we see that $a \Delta 0 = a$.

* THEOREM 30. $a \Delta e = a'$.

We have $a \vee e = e$ by *T 5 and $ae = a$ by *T 3; hence $a \Delta e$ is the unique solution of the equations $x \vee a = e$, $xe = 0$ in accordance with T 24 and D 4; since these equations have the solution a' by *P 5 and *D 2, we see that $a \Delta e = a'$.

* THEOREM 31. $a \Delta b = ab' \vee a'b$.

We have $a \Delta b = (a \vee b)(ab)' = (a \vee b)(a' \vee b') = aa' \vee ab' \vee a'b \vee bb' = 0 \vee ab' \vee a'b \vee 0 = ab' \vee a'b$ by *T 16₂, *T 12₁, P 3₁ — P 3₂ — T 13 — T 14, *L 3, P 1₁ — T 14 — P 4₁.

THEOREM 32. $(a \Delta b) \Delta (b \Delta c) = a \Delta c$.

Using *T 31, the commutative and associative laws for addition and multiplication, and the familiar properties of complementation, we have

$$\begin{aligned} (a \Delta b) \Delta (b \Delta c) &= (ab' \vee a'b)(b'c' \vee b'c)' \vee (ab' \vee a'b)'(b'c' \vee b'c) \\ &= (ab' \vee a'b)(b' \vee c)(b \vee c') \vee (a' \vee b)(a \vee b')(b'c' \vee b'c) \\ &= (ab' \vee a'b)(b'c' \vee bc) \vee (a'b' \vee ab)(b'c' \vee b'c) = ab'c' \vee a'bc \vee a'b'c \vee abc' \\ &= ac'(b' \vee b) \vee a'c(b \vee b') = ac'e \vee a'ce = ac' \vee a'c = a \Delta c. \end{aligned}$$

THEOREM 33. If $c = a \Delta b$, then $a = b \Delta c$; in particular $a = b \Delta (a \Delta b)$.

If $c = a \Delta b$, we have

$$a = a \Delta 0 = (a \Delta b) \Delta (b \Delta 0) = (a \Delta b) \Delta b = b \Delta (a \Delta b) = b \Delta c$$

by T 29, T 32, T 29 — T 25, T 26, H — T 25.

THEOREM 34. $a \Delta (b \Delta c) = (a \Delta b) \Delta c$.

We have $a \Delta (b \Delta c) = [a \Delta b] \Delta [b \Delta (b \Delta c)] = (a \Delta b) \Delta c$ by T 32, T 26 — T 33 — T 25.

THEOREM 35. In order that $a \Delta b = a \Delta c$ it is necessary and sufficient that $b = c$.

By T 28, $a \Delta b = a \Delta c$ if and only if $(a \Delta b) \Delta (a \Delta c) = 0$; the latter relation holds, by T 26, T 25, T 32, if and only if $b \Delta c = 0$; and, by T 28, $b \Delta c = 0$ if and only if $b = c$.

THEOREM 36. $ab = (a \vee b) \Delta (a \Delta b)$.

This result follows at once from T 27, T 26, T 33.

THEOREM 37. $a \vee b = (ab) \Delta (a \Delta b)$.

This result follows at once from T 27, T 26, T 33.

THEOREM 38. $(a \Delta b)c = (ac) \Delta (bc)$.

We have

$$\begin{aligned} (ac) \Delta (bc) &= (ac)(bc)' \vee (ac)'(bc) = (ac)(b' \vee c') \vee (a' \vee c')(bc) \\ &= (ab')c \vee (a'b)c = (ab' \vee a'b)c = (a \Delta b)c \end{aligned}$$

by T 31, and the familiar properties of addition, multiplication, and complementation.

THEOREM 39. $a \vee (a \Delta b) = a \vee b$.

We have

$$a \vee (a \Delta b) = (a \vee ab) \vee (a \Delta b) = a \vee [(a \Delta b) \vee ab] = a \vee (a \vee b) = (a \vee a) \vee b = a \vee b$$

by T 10, P 1₁ — T 14, D 4, T 14, P 6₁.

* THEOREM 40. $(a \Delta b)' = a' \Delta b = a \Delta b'$.

We have $(a \Delta b)' = (ab' \vee a'b)' = (a' \vee b)(a \vee b') = a'b' \vee ab = a' \Delta b$ by *T 31, and the familiar properties of addition, multiplication and complementation. By T 26, we have also $(a \Delta b)' = (b \Delta a)' = b' \Delta a = a \Delta b'$. We can also obtain this result from T 30, T 34, and T 26, in an obvious manner.

THEOREM 41. *If $a < c$ and $b < d$, then $a \Delta b < c \vee d$; in particular, $a \Delta b < a \vee b$.*

By T 18₁ and T 19, it is sufficient to show that $a \Delta b < a \vee b$. Now we have $(a \Delta b) \vee ab = a \vee b$ by D 4; and hence $(a \Delta b) < a \vee b$ by T 18₂ and T 20₂.

THEOREM 42. $a \Delta c < (a \Delta b) \vee (b \Delta c)$.

This result follows directly from T 32 by T 18₁, T 18₂, and T 41.

THEOREM 43. $(ab) \Delta (cd) < (a \Delta c) \vee (b \Delta d)$.

By suitable applications of *T 31 and the familiar properties of addition, multiplication, and complementation, we have

$$(ab) \Delta (cd) = (ab)(cd)' \vee (ab)'(cd) = (ac')b \vee (bd')a \vee (a'c)d \vee (b'd)c$$

and

$$(a \Delta c) \vee (b \Delta d) = ac' \vee bd' \vee a'c \vee b'd.$$

Then by using T 19 together with the relations $(ac')b < ac'$, $(bd')a < bd'$, $(a'c)d < a'c$, $(b'd)c < b'd$, which all follow from T 20₁, we obtain the result asserted in the statement of the theorem.

THEOREM 44. $(a \vee b) \Delta (c \vee d) < (a \Delta c) \vee (b \Delta d)$.

The proof is similar to that of the preceding theorem. We have

$$\begin{aligned} (a \vee b) \Delta (c \vee d) &= (a \vee b)(c \vee d)' \vee (a \vee b)'(c \vee d) \\ &= (ac')d' \vee (bd')c' \vee (a'c)b' \vee (b'd)a' < ac' \vee bd' \vee a'c \vee b'd = (a \Delta c) \vee (b \Delta d). \end{aligned}$$

We have now obtained the chief properties of the operation of differentiation. From an algebraic point of view, differentiation is the analogue in a Boolean algebra of both addition and subtraction in an abstract ring.⁸ The commutative and associative laws hold for differentiation, as we have shown in Theorems 26 and 34 respectively; and multiplication is distributive with respect to differentiation by Theorem 38. The connections between differentiation and the operations of addition, multiplication, and complementation are given by Theorems 37 and 39, Theorem 36, and *Theorems 31 and 40, respectively. In view of Theorems 25 and 37, we may regard a Boolean algebra as a system with double composition in which the given operations are those which we here call multiplication and differentiation, the operation which we here call addition being introduced through the equation $a \vee b = (ab) \Delta (a \Delta b)$. Alternatively, we may treat addition and differentiation as the given operations and introduce multiplication through the equation $ab = (a \vee b) \Delta (a \Delta b)$ of Theorem 36. We shall next call attention to an analogy which throws a good deal of light on Theorems 26, 28, and 42. If we call $a \Delta b$ the Boolean distance from the element a to the element b , we find that these theorems simply reflect familiar properties of ordinary numerical distances: Theorems 17 and 25 show that the Boolean distance is a function of the arguments a and b defined and single-valued for all a and b in a Boolean algebra A , with

⁸ See the remarks in the introduction.

values greater than (or equal to) 0 belonging to A ; Theorem 26 shows that the Boolean distance is symmetric in its two arguments; Theorem 28 shows that the Boolean distance is equal to 0 if and only if $a = b$; Theorem 42 shows that the Boolean distance satisfies the "triangle inequality"; and Theorem 38 shows that the Boolean distance is a homogeneous function of degree one under Boolean multiplication of its arguments. Finally, we point out that the reciprocity established in Theorem 33 can be brought out in another way which is interesting in itself but not important later. Let us define a triadic relation by writing $R(a, b, c)$ to indicate that the equations $a(b \vee c) = a$, $b(c \vee a) = b$, $c(a \vee b) = c$, $abc = 0$ hold for the elements a, b, c . It is evident that this relation is symmetric in the related elements—in other words, that the relation persists under any permutation of the elements a, b, c . We can now show that, when b and c are given, there exists a unique element a such that $R(a, b, c)$ —namely, the element $a = b \Delta c$: for $R(a, b, c)$ implies $a \vee bc = (b \vee c)a \vee bc = (ba \vee bc) \vee (ac \vee bc) = b \vee c$, $a(bc) = 0$, and hence $a = b \Delta c$; and $a = b \Delta c$ implies

$$\begin{aligned} a(b \vee c) &= (b \Delta c)[bc \Delta (b \Delta c)] = (b \Delta c)bc \Delta (b \Delta c) = 0 \Delta a = a, \\ b(c \vee a) &= b[c \vee (b \Delta c)] = b(b \vee c) = b, \\ c(a \vee b) &= c[(b \Delta c) \vee b] = c(c \vee b) = c, \quad abc = (b \Delta c)bc = 0, \end{aligned}$$

and hence $R(a, b, c)$. Thus the symmetry of our triadic relation is reflected in the reciprocity of Theorem 33.

2. *Connections with the Postulates of Huntington and De Morgan.* We must now show that our definition of Boolean algebras is essentially equivalent to at least one of the standard definitions. Naturally it is convenient to make the necessary comparison with some definition which, like ours, is based upon a system with double composition. The two best-known sets of postulates of this type are those due respectively to Huntington and to De Morgan. We repeat them here:⁹

HUNTINGTON'S POSTULATES. *A Boolean algebra A is a system with double composition which satisfies the postulates*

Postulate 1₁. $a \vee b = b \vee a$; Postulate 1₂. $ab = ba$;

Postulate 3₁. $a(b \vee c) = ab \vee ac$;

Postulate 3₂. $a \vee bc = (a \vee b)(a \vee c)$;

⁹ See Huntington, *Transactions of the American Mathematical Society*, vol. 5 (1904), pp. 288-309; De Morgan, *Algebra della Logica*, Naples, 1907; B. A. Bernstein, *University of California Publications in Mathematics*, vol. 1 (1914), pp. 87-96.

Postulate 4₁. There exists an element 0 in A such that $a \vee 0 = a$ for every element a in A;

Postulate 4₂. There exists an element e in A such that $ea = a$ for every element a in A;

Postulate 5₃. If the elements 0 and e of Postulates 4₁ and 4₂ respectively exist and are unique, then the simultaneous equations $x \vee a = e$, $xa = 0$ have a solution for every element a in A;

Postulate 7. There exist elements a and b in A such that $a \neq b$.

DEL RE'S POSTULATES. A Boolean algebra A is a system with double composition which satisfies the postulates

Postulates 1₁, 1₂, 3₁, 4₁, 4₂, 5₃, 6₁, 7; [Postulate 2₁. $(a \vee b) \vee c = a \vee (b \vee c)$.]

We have, of course, numbered the various postulates concerned in such a way as to facilitate comparison with our own system. Furthermore, we have made some minor changes in phraseology in order to eliminate trivial differences in statement.

We now observe that, of the postulates which occur in Huntington's and Del Re's systems, every one except Postulate 7 either occurs in our system or is deducible therefrom. We have established Postulate 1₂ as Theorem 13 in our system, Postulate 2₁ as Theorem 14, Postulate 3₃ as Theorem 11, Postulate 4₂ as Theorem 3; and we can establish Postulate 5₃ as a consequence of Postulate 5 and Theorem 3.

We next observe that, of the postulates which occur in our system, every one either occurs in Huntington's system or is deducible therefrom without the use of Postulate 7. We can establish Postulate 3₂ in Huntington's system since $(a \vee b)c = c(a \vee b) = ca \vee cb = ac \vee bc$ by P 1₂, P 3₁, P 1₂ — P 1₂. We can establish Postulate 5 as follows: in Huntington's system we establish the existence and uniqueness of the element 0 by combining Postulates 1₁ and 4₁ just as we did in proving Theorem 1 in our system; we establish the existence and uniqueness of the element e by combining Postulates 1₂ and 4₂ in a similar manner; we then use Postulate 5₃ to show that, for the indicated elements 0 and e, the simultaneous equations $x \vee a = e$, $xa = 0$ have a solution, whatever the element a in A. We establish Postulate 6₁ from the relations $a = a \vee 0 = a \vee xa = (a \vee x)(a \vee a) = (x \vee a)(a \vee a) = e(a \vee a) = a \vee a$ by P 4₁, P 5₃, P 3₃, P 1₁, P 5₃, P 4₂, the element x being chosen as a solution of the simultaneous equations $x \vee a = e$, $xa = 0$, in accordance with Postulate 5₃ and the previous remarks concerning the existence and uniqueness of the elements 0 and e. Finally, we establish Postulate 6₂ from the relations

$a = ea = (x \vee a)a = xa \vee aa = 0 \vee aa = aa \vee 0 = aa$ by P 4₂, P 5₃, P 3₁, P 5₃, P 1₁, P 4₁, the element x being chosen as in the preceding demonstration.

We observe finally that, of the postulates which occur in our system, every one either occurs in Del Re's system or is deducible therefrom without the use of Postulates 2₁ and 7. In fact, we see that Postulates 3₂, 5, and 6₂ can be established in Del Re's system by arguments precisely the same as those which were effective in Huntington's. *Since Postulate 2₁ can now be deduced as Theorem 14 from the postulates of our system, it follows that this postulate is redundant in Del Re's system.*¹⁰

Postulate 7 plays no part in deductions from the systems of Huntington and Del Re; indeed, it serves only to exclude the essentially trivial case of an algebra with exactly one element. It is largely a matter of taste whether one shall or shall not reject this case. We find it convenient to consider one-element algebras and hence to omit Postulate 7 from our list.

We may summarize the foregoing comments as follows:

THEOREM 45. *Definition 1 is equivalent to Huntington's definition of Boolean algebras and also to Del Re's, provided that Postulate 7 is suppressed from their lists of postulates:*

3. *Postulates for generalized Boolean algebras.* It is of some interest to examine the possibility of introducing a system with double composition which possesses most of the peculiar properties of Boolean algebras without containing a unit element. The analogy with the theory of abstract rings suggests that the direct or indirect postulation of the unit element should be avoided; and examples from the theory of classes—for instance, the example of all finite subclasses of a given infinite class, or the example of all Lebesgue or Borel measurable sets of finite measure in n -space—indicate the existence of interesting algebras without unit. We shall therefore introduce the following definition:

DEFINITION 5. *A generalized Boolean algebra A is a system with double composition which satisfies the further postulates:*

¹⁰ The example which has been given by B. A. Bernstein, *University of California Publications in Mathematics*, vol. 1 (1914), pp. 87-96, especially p. 95, to show the independence of Postulate 2₁ in Del Re's system is known to be irrelevant, as Professor Bernstein informed the writer, first by word of mouth and then by letter when consulted on the situation presented here: for both Postulate 2₁ and Postulate 3₁ fail to hold in this example, as we see from the relations

$$(e_1 \vee e_2) \vee e_3 = e_{12} \vee e_3 = e_3 \neq e_{123} = e_1 \vee e_{23} = e_1 \vee (e_2 \vee e_3)$$

and

$$e_2(e_{12} \vee e_3) = e_2e_3 = e_0 \neq e_2 = e_2 \vee e_0 = e_2e_{12} \vee e_2e_3.$$

Postulate 1₁. $a \vee b = b \vee a$;

Postulate 2₂. $a(bc) = (ab)c$;

Postulate 3₁. $a(b \vee c) = ab \vee ac$;

Postulate 4₁. There exists an element 0 in A such that $a \vee 0 = a$ for every element a in A ;

Postulate 5₁. If there exists an element 0 with the property required by Postulate 4₁ and if a and b are elements of A such that $ba = a$, then there exists at least one such element 0, independent of a and b , for which the simultaneous equations $x \vee a = b$, $xa = 0$ have a solution in A ;

Postulate 5₂. If there exists an element 0 with the property required by Postulate 4₁ and if a and b are elements of A such that $ab = a$, then there exists at least one such element 0, independent of a and b , for which the simultaneous equations $x \vee a = b$, $ax = 0$ have a solution in A ;

Postulate 6₁. $a \vee a = a$;

Postulate 6₂. $aa = a$.

Any element such as that postulated in Postulate 4₁ is called a zero.

The list of postulates given in this definition is obtained by certain obvious modifications of the list given in Definition 1. We begin by replacing * Postulate 5 by the more complicated, symmetrically related Postulates 5₁ and 5₂, in order to obtain a system without postulated unit in which the indicated simultaneous equations still possess solutions; we may point out that, in the presence of the remaining postulates of Definition 1, the relations $ba = a$ and $ab = a$ of the new postulates are necessary conditions for the existence of such solutions. This substitution unfortunately renders the list of postulates inadequate for our purpose and therefore forces us to proceed further. We elect to add Postulate 2₂ at this stage. We find, however, that the addition makes each of the Postulates 3₁, 3₂ redundant in the presence of the others¹¹; but we are able to bring our series of changes to a close with the suppression of one or the other of these two postulates. We have chosen to postulate the distributive law for multiplication on the left; but, if we had chosen to postulate the distributive law for multiplication on the right, we could have recovered the list of postulates given in Definition 5 simply by introducing a new operation \circ defined by the relation $a \circ b = ba$. It may be noted that the

¹¹ I am indebted to Professor A. A. Bennett for this observation; the argument given in Theorems 47-50 below is due essentially to him.

finite subclasses of a fixed infinite class constitute a system satisfying the postulates of Definition 5 but possessing no unit.

By way of justifying the terminology introduced in Definition 5, we may note the following result:

THEOREM 46. *Every Boolean algebra is a generalized Boolean algebra.*

We have only to prove that any system with double composition which satisfies the postulates of Definition 1 satisfies the additional postulates occurring in Definition 5. Now Postulate 2_2 has been established as Theorem 15, Postulate 5_1 has been established as Theorem 16₂, and Postulate 5_2 follows from Postulate 5_1 by an application of Theorem 13.

In discussing generalized Boolean algebras, it is our aim to show that all the unstarred theorems of § 1 remain valid in these more general systems. We shall condense the necessary developments as much as possible, using the form of presentation adopted in § 1.

THEOREM 47. *The element 0 of Postulate 4₁ exists and is unique.*

By P 4₁, such an element exists. If 0_1 and 0_2 are such elements, then $0_1 = 0_1 \vee 0_2 = 0_2 \vee 0_1 = 0_2$ by P 4₁, P 1₁, P 4₁.

We shall use Theorem 47 without specific reference in the sequel.

THEOREM 48₁. $0a = 0$.

By P 6₂ — P 5₁, there exists an element x with the property X that $x \vee a = a$, $xa = 0$. We now have $x = x \vee 0 = xx \vee xa = x(x \vee a) = xa = 0$ by P 4₁, P 6₂ — X, P 3₁, X, X. Combining this result with X, we have $0a = xa = 0$.

THEOREM 48₂. $a0 = 0$.

By P 6₂ — P 5₂, there exists an element x with the property X that $x \vee a = a$, $ax = 0$. Hence we have $a0 = a(ax) = (aa)x = ax = 0$ by X, P 2₂, P 6₂, X.

LEMMA 11. $ab = 0$ implies $ba = 0$.

We have $ba = (ba)(ba) = b[a(ba)] = b[(ab)a] = b[0a] = b0 = 0$ by P 6₂, P 2₂, P 2₂, H, T 48₁, T 48₂.

LEMMA 12. $(ab)a = ab$.

We have $ab = (aa)b = a(ab)$ by P 6₂, P 2₂. Hence by P 5₁ there exists an element x with the property X₁ that $x \vee ab = a$, $x(ab) = 0$; and, by

X_1 — L 11, this element x has also the property X_2 that $(ab)x = 0$. Thus we have $(ab)a = (ab)(x \vee ab) = (ab)x \vee (ab)(ab) = 0 \vee ab = ab \vee 0 = ab$ by X_1 , P 3₁, X_2 — P 6₂, P 1₁, P 4₁.

THEOREM 49. $ab = ba$.

We have $(ab)(ba) = [(ab)b]a = [a(bb)]a = [ab]a = ab$ by P 2₂, P 2₂, P 6₂, L 12. Hence by P 5₂ there exists an element x with the property X_1 that $x \vee ab = ba$, $(ab)x = 0$. By X_1 — L 11 this element has the property X_2 that $x(ab) = 0$. It has also the property X_3 that $x(ba) = 0$, since

$$(ba)x = [(ba)b]x = b[(ab)x] = b0 = 0$$

by L 12, P 2₂ — P 2₂, X_1 , T 48₂, and L 11 is therefore applicable. Finally, it has the property X_4 that $x = 0$, since

$$x = x \vee 0 = xx \vee x(ab) = x(x \vee ab) = x(ba) = 0$$

by P 4₁, P 6₂ — X_2 , P 3₁, X_1 , X_3 . Hence we conclude that

$$ab = ab \vee 0 = ab \vee x = x \vee ab = ba$$

by P 4₁, X_4 , P 1₁, X_1 .

THEOREM 50. $(a \vee b)c = ac \vee bc$.

We have $(a \vee b)c = c(a \vee b) = ca \vee cb = ac \vee bc$ by T 49, P 3₁, T 49 — T 49; in other words, Postulate 3₂ is valid in a generalized Boolean algebra.

If a is an arbitrary element of A , we may consider the class of all elements c such that $c = ab$ for some element b in A . We shall denote this class by the symbol $\mathfrak{a}(a)$.

THEOREM 51. *If a is an arbitrary element in the generalized Boolean algebra A , then the class $\mathfrak{a}(a)$ is a subsystem of A and is in addition a Boolean algebra, its zero being the element 0 in A and its unit being the element a .*

If c_1 and c_2 are elements of $\mathfrak{a}(a)$, we can write $c_1 = ab_1$, $c_2 = ab_2$; thus, $c_1 \vee c_2 = ab_1 \vee ab_2 = a(b_1 \vee b_2)$ by P 3₁ and $c_1 c_2 = (ab_1)(ab_2) = a[b_1(ab_2)]$ by P 2₂; and hence $c_1 \vee c_2$ and $c_1 c_2$ are elements of $\mathfrak{a}(a)$. Thus we see that $\mathfrak{a}(a)$ is a subsystem of the system A . We now note that 0 and a are both elements of $\mathfrak{a}(a)$, since we have $0 = a0$ by T 48₂ and $a = aa$ by P 6₂. Now, since $\mathfrak{a}(a)$ is a subsystem of A containing the element 0 and since Postulates 1₁, 3₁, 3₂ (by T 50), 4₁, 6₁, and 6₂ are valid in A , these postulates are also valid in

$\mathfrak{a}(a)$. Hence, to prove that $\mathfrak{a}(a)$ is a Boolean algebra, we have only to establish Postulate 5 for this system. This may be done as follows: we first note that if c is any element in $\mathfrak{a}(a)$ we can write $c = ab$ and therefore have $ac = a(ab) = (aa)b = ab = c$ by H, P 2₂, P 6₂, H; hence, by P 5₁, there exists an element x in A with the property X that $x \vee c = a$, $xc = 0$; and, since $ax = xa = x(x \vee c) = xx \vee xc = x \vee 0 = x$ by T 49, X, P 3₁, P 6₂ — X, P 4₁, this element x , as well as the elements 0, a , and c , belongs to $\mathfrak{a}(a)$. In particular, we see from the foregoing results that the zero in $\mathfrak{a}(a)$ is the element 0 in A and that the unit in $\mathfrak{a}(a)$ is the element a .

THEOREM 52. *If c is an element of $\mathfrak{a}(a)$, then $\mathfrak{a}(c) \subset \mathfrak{a}(a)$.*

If d is an element in $\mathfrak{a}(c)$, we can write $d = cb_1$ and $c = ab_2$ and thus find that $d = cb_1 = (ab_1)b_2 = a(b_1b_2)$ by H, H, P 2₂. Hence d is in $\mathfrak{a}(a)$ and $\mathfrak{a}(c) \subset \mathfrak{a}(a)$.

THEOREM 53. *The class $\mathfrak{a}(a \vee b)$ contains the elements a and b .*

Since a and ab are elements of the Boolean algebra $\mathfrak{a}(a)$ in accordance with T 51, we have $a \vee ab = a$ by T 10 and hence

$$a = a \vee ab = aa \vee ab = a(a \vee b) = (a \vee b)a$$

by P 6₂, P 3₁, T 49; it follows that a is an element of $\mathfrak{a}(a \vee b)$. Since b and ba are elements of the Boolean algebra $\mathfrak{a}(b)$, we have similarly

$$b = b \vee ba = ba \vee b = ba \vee bb = b(a \vee b) = (a \vee b)b$$

by T 10, P 1₁, P 6₂, P 3₁, T 49 and hence conclude that b is an element of $\mathfrak{a}(a \vee b)$.

THEOREM 54. *The class $\mathfrak{a}(a \vee [b \vee (c \vee d)])$ contains the elements a, b, c, d .*

By T 53 we see that $\mathfrak{a}(a \vee [b \vee (c \vee d)])$ contains a and $b \vee (c \vee d)$, $\mathfrak{a}(b \vee (c \vee d))$ contains b and $c \vee d$, and $\mathfrak{a}(c \vee d)$ contains c and d . By T 52 we see that $\mathfrak{a}(c \vee d) \subset \mathfrak{a}(b \vee (c \vee d)) \subset \mathfrak{a}(a \vee [b \vee (c \vee d)])$. Hence the class $\mathfrak{a}(a \vee [b \vee (c \vee d)])$ contains the elements a, b, c, d .

THEOREM 55. *All unstarred propositions and definitions of § 1 are valid in a generalized Boolean algebra, provided that the term "Boolean algebra" is supplanted wherever it occurs by the term "generalized Boolean algebra."*

A brief survey of the definitions and propositions of § 1 shows that the starred definitions and propositions are precisely those involving the unit element e and the related operation of complementation. Such definitions and theorems obviously cannot be carried over to a generalized Boolean algebra.

To assert that the remaining definitions and propositions of § 1 can be carried over to a generalized Boolean algebra therefore means that the operations of addition, multiplication, and differentiation, and the relation "less than" are defined in such an algebra and are subject to the same formal rules as in a Boolean algebra.

In verifying this assertion, we must examine in succession the various unstarred definitions and propositions of § 1. In deciding whether or not a definition or proposition remains valid in a generalized Boolean algebra, we may appeal to several different criteria. A definition can be retained whenever it is stated directly in terms of the fundamental postulated operations of addition and/or multiplication or indirectly through preceding unstarred definitions and/or propositions. This criterion applies at once to Definitions 3, 4; of course, Definition 4 reposes upon the preceding unstarred Theorem 24. A proposition which is stated in terms of the fundamental postulated operations and/or preceding unstarred definitions remains valid (1) whenever it is one of those postulated or proved in the present section; (2) whenever it reposes only upon preceding unstarred propositions and/or starred propositions which have adequate substitutes valid in a generalized Boolean algebra; (3) whenever it asserts a relation which can be treated by regarding the related elements as belonging to a Boolean algebra constructed in accordance with Theorems 51 and 54 of the present section. We apply criterion (1) in the following instances: Postulates 1_1 , 3_1 , 4_1 , 6_1 , 6_2 of Definition 1 appear as postulates in Definition 5; Postulate 3_2 has been proved as Theorem 50 of the present section; Theorem 1 appears essentially as Theorem 47 of the present section; Theorem 6 appears, subdivided as Theorems 48_1 and 48_2 of the present section; Theorem 13 appears as Theorem 49 of the present section and Theorem 15 appears as Postulate 2_2 of Definition 5. We can apply criterion (2) above in the following instances: Theorems 16_1 , 17 , 18_1 , 18_2 , 19 , 20_1 , 20_2 , 21 , 24 , 25 , 26 , 27 , 28 , 29 , 33 , 34 , 35 , 36 , 37 , 39 , 41 , 42 , and Lemmas 8_1 , 8_2 , 10 . In the case of Theorem 24, the proof involves the starred Theorem 16_2 , which is used, however, merely to establish the existence of a solution of the simultaneous equations under consideration; but in a generalized Boolean algebra the existence of the desired solution is established by Postulates 5_1 and 5_2 (which we may now regard as equivalent by virtue of the commutative law for multiplication). Finally we apply criterion (3) in the following instances: Theorems 9, 10, 11, 14, 32, 43, and 44. We shall content ourselves with giving detailed discussions of two typical cases. First let us consider the demonstration of Theorem 9: if 0 is the zero element of a generalized Boolean algebra A and a and b are arbitrary elements in A with the

property that $a \vee b = 0$, we may regard $0, a, b$ as elements of the Boolean algebra $\mathfrak{A}(a \vee [b \vee (c \vee d)])$ in accordance with Theorems 51 and 54, the element 0 being the zero of this algebra; Theorem 9 is now applicable within this Boolean algebra and shows that $a = b = 0$, as desired. Let us also consider the demonstration of Theorem 32: whether we regard the elements a, b, c and the operation Δ as belonging to the generalized Boolean algebra A or to the Boolean algebra $\mathfrak{A}(a \vee [b \vee (c \vee d)])$ contained in A , the elements $a \Delta b, b \Delta c, a \Delta c$ are obtained as the unique solutions of certain fixed pairs of simultaneous equations and are therefore independent of the interpretation chosen; now under the second interpretation, and hence also under the first, these elements belong to the algebra $\mathfrak{A}(a \vee [b \vee (c \vee d)])$; in the same way we find that the element $(a \Delta b) \Delta (b \Delta c)$ also belongs to this algebra; thus we can apply Theorem 32 within the Boolean algebra $\mathfrak{A}(a \vee [b \vee (c \vee d)])$ and conclude that $(a \Delta b) \Delta (b \Delta c) = a \Delta c$, as we wished to do. In summary, we may say that the three indicated criteria cover all the unstarred propositions of § 1 and thus lead to the present theorem.

We shall conclude the present section with a brief discussion of certain conditions under which a generalized Boolean algebra is a Boolean algebra. We have first:

THEOREM 56. *A necessary and sufficient condition that a generalized Boolean algebra A be a Boolean algebra is that it contain an element e with one or both of the equivalent properties that $ae = a, ea = a$ whatever the element a in A . If such an element exists, it is the unit of the Boolean algebra A .*

If A is a Boolean algebra, such an element exists by Postulates 4₁, 5 and Theorems 2, 3. If such an element exists, then Postulates 5₁, 5₂ show that the simultaneous equations $x \vee a = e, xa = 0$ have a solution whatever the element a in A ; thus the system A satisfies Postulate 5 as well as all the unstarred postulates of Definition 1, and is therefore a Boolean algebra with the element e as its unit.

THEOREM 57. *If A is a generalized Boolean algebra with finite cardinal number n , then A is a Boolean algebra.*

By hypothesis, there exist elements a_1, \dots, a_n such that $a_j \neq a_k$ for $j \neq k, j, k = 1, \dots, n$ while $a \in A$ implies $a = a_k$ for some integer $k, k = 1, \dots, n$. Now the element $e = a_1 \vee \dots \vee a_n$ has the property that $ae = a$ whatever the element a in A : for, putting $a = a_k$, we have

$$ae = a_k(a_1 \vee \dots \vee a_n) = a_k a_k \vee a_k(a_1 \vee \dots \vee a_{k-1} \vee a_{k+1} \vee \dots \vee a_n) = a_k = a$$

by H, P 1₁ — P 1.3₁ — T 14 — T 55, P 6₂ — T 10 — T 55, H. Thus A is a Boolean algebra in accordance with Theorem 56.

4. *Independence of the postulates.* The consistency of the sets of postulates given in Definitions 1 and 5 for Boolean algebras and generalized Boolean algebras respectively is shown by exhibiting specific systems with double composition in which all the listed postulates are valid. The system with a single element 0 where $0 \vee 0$ and 00 are both 0 is one such system; and the system with two elements 0 and e , where $0 \neq e$, where $0 \vee 0$, 00 , $0e$, $e0$ are all 0, and where $0 \vee e$, $e \vee 0$, $e \vee e$, ee are all e , is another. As we have already pointed out by means of examples, there are systems with double composition which are generalized Boolean algebras without being Boolean algebras; in view of Theorem 57, such systems are necessarily infinite.

In the following examples, we shall denote by $*KP\alpha$ a system with double composition in which Postulate α of Definition 1 is not satisfied while the remaining postulates of that definition are valid; and we shall denote by $KP\alpha$ a system with double composition in which Postulate α of Definition 5 is not satisfied while the remaining postulates of that definition are valid. In each example, the system considered is finite and its elements are designated by appropriate letters, different letters always denoting unequal elements and the letters 0 and e being reserved for elements with the properties described in Postulates 4₁ and 5; and the fundamental operations are defined by an addition table placed at the left and a multiplication table placed at the right, the tables being arranged so that $a \vee b$ or ab appears in the row labeled " a " and the column labeled " b ." In many cases we shall point out various properties of the systems exhibited beyond those mentioned above. In those instances where we assert the failure of a certain proposition in a given system, we shall indicate at least one specific case of the asserted failure; but in those instances where we assert the validity of a certain proposition in a given system, we shall merely report the result of checking all possible cases.

$*KP 1_1, KP 1_1.$

	0 ₁	0 ₂		0 ₁	0 ₂
0 ₁	0 ₁	0 ₁	0 ₁	0 ₁	0 ₁
0 ₂	0 ₂	0 ₂	0 ₂	0 ₁	0 ₂

It is evident that P 1₁ fails here, since $0_1 \vee 0_2 = 0_1 \neq 0_2 = 0_2 \vee 0_1$. P 2₂, P 3₁, P 3₂, P 6₁, P 6₂ are easily verified; and it is also clear that P 4₁ is valid whichever of the two elements 0₁, 0₂ is designated as the element 0 of that

postulate. If we choose 0_1 as the element 0 and also as the element e of *P 5, we find that *P 5 is satisfied: for the simultaneous equations $x \vee a = 0_1$, $xa = 0_1$ have the solution 0_1 , whatever the element a .

Neither P 5₁ nor P 5₂ is valid here: for none of the four sets of simultaneous equations

$$\begin{array}{ll} x \vee 0_2 = 0_2, & x0_2 = 0_1; \\ x \vee 0_1 = 0_2, & x0_1 = 0_2; \end{array} \quad \begin{array}{ll} x \vee 0_2 = 0_2, & 0_2x = 0_1; \\ x \vee 0_1 = 0_2, & 0_1x = 0_2; \end{array}$$

has a solution although the relations $0_20_2 = 0_2$, $0_10_2 = 0_20_1 = 0_1$ are valid. It is easily verified that Theorems 10, 11, 13, 14, 15 hold in this system.

*KP 3 ₁ .	0	a	b	c	e		0	a	b	c	e
0	0	a	b	c	e		0	0	0	0	0
a	a	a	e	e	e		a	0	a	0	a
b	b	e	b	b	e		b	0	0	b	c
c	c	e	b	c	e		c	0	0	b	c
e	e	e	e	e	e		e	0	a	b	c

The failure of P 3₁ is exhibited by the relations

$$c(a \vee b) = ce = c \neq b = 0 \vee b = ca \vee cb.$$

It can be verified directly that P 1₁, P 3₂, P 4₁, *P 5, P 6₁, P 6₂ are valid in this system; and also that P 2₂, P 5₂, T 9, T 14, T 15 are valid here. On the other hand, P 5₁ fails since $bc = c$ while the simultaneous equations $x \vee c = b$, $xc = 0$ have no solution; T 10 fails since $c \vee cb = c \vee b = b \neq c$; T 11 fails since $(c \vee a)(c \vee b) = eb = b \neq c = c \vee 0 = c \vee ab$; and T 13 fails since $bc = c \neq b = cb$.

*KP 3 ₂ .	0	a	e		0	a	e
0	0	a	e		0	0	0
a	a	a	e		a	a	a
e	e	e	e		e	0	e

The failure of P 3₂ is evident from the relations

$$(e \vee a)0 = e0 = 0 \neq a = 0 \vee a = e0 \vee a0.$$

It can be verified directly that P 1₁, P 3₁, P 4₁, *P 5, P 6₁, P 6₂ are valid in this system; and also that P 2₂, P 5₁, T 9, T 14, T 15 are valid here. On the

other hand, $P 5_2$ fails since $a0 = a$ while the equations $x \vee a = 0$, $ax = 0$ have no solution; $T 10$ fails since $0 \vee a0 = 0 \vee a = a \neq 0$; $T 11$ fails since

$$(a \vee e)(a \vee 0) = ea = 0 \neq a = a \vee 0 = a \vee e0;$$

and $T 13$ fails since $a0 = a \neq 0 = 0a$.

*KP 4₁, KP 4₁.

	a	b	c		a	b	c
a	a	c	c	a	a	a	a
b	c	b	c	b	a	b	c
c	c	c	c	c	a	c	c

The failure of $P 4_1$ is evident from the addition table. Since the element 0 described in $P 4_1$ does not exist, $*P 5$, $P 5_1$, and $P 5_2$ are satisfied vacuously. It can be verified directly that $P 1_1$, $P 2_2$, $P 3_1$, $P 3_2$, $P 6_1$, $P 6_2$ are valid in this system; and also that $T 11$, $T 13$, $T 14$, $T 15$ are valid here. On the other hand, $T 10$ fails since $b \vee ba = b \vee a = c \neq b$.

*KP 5.

	0	a		0	a
0	0	a	0	0	a
a	a	a	a	a	a

The failure of $*P 5$ is evident from the fact that even the single equation $xa = 0$ has no solution in this system. It is easily verified that $P 1_1$, $P 3_1$, $P 3_2$, $P 4_1$, $P 6_1$, $P 6_2$ are valid here; and that $P 2_2$, $T 9$, $T 11$, $T 13$, $T 14$, $T 15$ are also valid. On the other hand, $P 5_1$ fails since $0a = c$ while the equations $x \vee a = 0$, $xa = 0$ have no solution; $P 5_2$ fails since $a0 = a$ while the equations $x \vee a = 0$, $ax = 0$ have no solution; and $T 10$ fails since $0 \vee 0a = 0 \vee a = a \neq 0$.

*KP 6₁, KP 6₁.

	0	e		0	e
0	0	e	0	0	0
e	e	0	e	0	e

The failure of $P 6_1$ is evident from the relation $e \vee e = 0$. It is easily verified that $P 1_1$, $P 2_2$, $P 3_1$, $P 3_2$, $P 4_1$, $*P 5$, $P 5_1$, $P 5_2$, $P 6_2$ are valid in this system; and also that $T 13$, $T 14$, $T 15$ are valid here. On the other hand, $T 9$ fails since $e \vee e = 0$; $T 10$ fails since $e \vee ee = e \vee e = 0 \neq e$; and $T 11$ fails since $(e \vee e)(e \vee 0) = 0e = 0 \neq e = e \vee 0 = e \vee e0$. This system is

of particular interest because it is a ring, isomorphic to the ring of integers taken modulo 2.

*KP 6₂, KP 6₂.

	0	e
0	0	e
e	e	e

	0	e
0	0	0
e	0	0

The failure of P 6₂ is evident from the relation $ee = 0$. It is easily verified that P 1₁, P 2₂, P 3₁, P 3₂, P 4₁, *P 5, P 5₁, P 5₂, P 6₁ are valid in this system; and also that T 9, T 10, T 13, T 14, T 15 are valid here. On the other hand, T 11 fails since $(e \vee 0)(e \vee 0) = ee = 0 \neq e = e \vee 0 = e \vee 00$.

KP 2₂.

	0	a	b	c
0	0	a	b	c
a	a	a	c	b
b	b	c	b	a
c	c	b	a	c

	0	a	b	c
0	0	0	0	0
a	0	a	c	b
b	0	c	b	a
c	0	b	a	c

It is evident that P 2₂ fails in this system since

$$a(bc) = aa = a \neq c = cc = (ab)c.$$

It is easily verified that P 1₁, P 3₁, P 4₁, P 5₁, P 5₂, P 6₁, P 6₂ are valid here; and that P 3₂, T 9, T 13 are also valid. On the other hand, *P 5 fails since the relations $x \vee y = z$, $xy = 0$, imply $x = 0$, $z = y$ or $y = 0$, $z = x$ and hence preclude the existence of the required element e ; T 10 fails since $a \vee ab = a \vee c = b \neq a$; T 11 fails $(a \vee 0)(a \vee b) = ac = b \neq a = a \vee 0 = a \vee 0b$; T 14 fails since $a \vee (b \vee b) = a \vee b = c \neq a = c \vee b = (a \vee b) \vee b$; and T 15 fails since it is equivalent to P 2₂.

KP 3₁.

	0	a	b	c	e
0	0	a	b	c	e
a	a	a	e	e	e
b	b	e	b	e	e
c	c	e	e	c	e
e	e	e	e	e	e

	0	a	b	c	e
0	0	0	0	0	0
a	0	a	0	0	a
b	0	0	b	0	b
c	0	0	0	c	c
e	0	a	b	c	e

The failure of P 3₁ is evident from the relations

$$a(b \vee c) = ae = a \neq 0 = 0 \vee 0 = ab \vee ac.$$

It is easily verified that $P 1_1, P 2_2, P 4_1, P 5_1, P 5_2, P 6_1, P 6_2$ are valid in this system; and also that $*P 5, T 9, T 10, T 13, T 14, T 15$ are valid here. On the other hand, $T 11$ fails since $(a \vee b)(a \vee c) = ee = e \neq a = a \vee 0 = a \vee bc$. This particular system may be interpreted in the following way: e is a fixed two-dimensional linear manifold; 0 is a fixed 0-dimensional linear manifold in e ; a, b, c are three distinct 1-dimensional linear manifolds in e containing 0 ; the sum of any two of these linear manifolds is the least linear manifold containing both, and the product is the largest linear manifold contained in both.

KP 5₁.

	0	a		0	a
0	0	a	0	0	a
a	a	a	a	0	a

The failure of $P 5_1$ is evident from the fact that $0a = a$ while the simultaneous equations $x \vee a = 0, xa = 0$ have no solution. It is easily verified that $P 1_1, P 2_2, P 3_1, P 4_1, P 5_2, P 6_1, P 6_2$ are valid in this system; and also that $P 3_2, T 9, T 11, T 14, T 15$ are valid here. On the other hand, $*P 5$ fails since neither the equations $x \vee a = 0, xa = 0$ nor the equations $x \vee a = a, xa = 0$ have a solution and consequently no element with the properties demanded of e exists; $T 10$ fails since $0 \vee 0a = 0 \vee a = a \neq 0$; and $T 13$ fails since $0a = a \neq 0 = a0$.

KP 5₂.

	0	a		0	a
0	0	a	0	0	0
a	a	a	a	a	a

The failure of $P 5_2$ is evident from the fact that $a0 = a$ while the simultaneous equations $x \vee a = 0, ax = 0$ have no solution. It is easily verified that $P 1_1, P 2_2, P 3_1, P 4_1, P 5_1, P 6_1, P 6_2$ are valid in this system; and also that $P 3_2, T 9, T 11, T 14, T 15$ are valid here. On the other hand, $*P 5$ fails since neither the equations $x \vee 0 = a, x0 = 0$ nor the equations $x \vee a = 0, xa = 0$ have a solution and consequently no element with the properties demanded of e exists; $T 10$ fails since $0 \vee a0 = 0 \vee a = a \neq 0$; and $T 13$ fails since $0a = 0 \neq a = a0$. The system under consideration is obtained from KP 5₁ simply by reflecting the multiplication table in its principal diagonal, so that the product xy in KP 5₂ is the same as the product yx in KP 5₁.

In conclusion, we see that each of the various postulates in Definition 1 or Definition 5 is independent of the remaining postulates of that definition.

ON FINITE BOOLEAN ALGEBRAS.¹

By B. A. BERNSTEIN.

Introduction. In every finite Boolean algebra there exists a set of elements, which I shall call the "minimals" of the algebra, that play a rôle in the algebra very much like that played by the prime numbers in arithmetic. Huntington² introduced these elements, under the term "irreducible,"³ for the purpose of proving the order theorem for finite Boolean algebras. The object of my paper is to discuss the minimal elements more fully and to make further application of the properties of these elements. The application will be to some questions concerning sub-algebras of a finite Boolean algebra, to a problem in dichotomy, and to arithmetic representations of finite Boolean algebras. Incidentally, Huntington's results concerning the minimal elements will be obtained from a different starting point. Incidentally, also, a proof of the order theorem for finite Boolean algebras will be obtained without the help of the minimal elements.⁴

1. *The minimals.* Consider any Boolean algebra B . If a_1, a_2, \dots, a_n be elements of B , and 1 the universe element, then

$$(i) \quad 1 = (a_1 + a'_1)(a_2 + a'_2) \cdots (a_n + a'_n)$$

$$(ii) \quad = a_1 a_2 \cdots a_n + a_1 a_2 \cdots a'_n + \cdots + a'_1 a'_2 \cdots a'_n.$$

The expression (ii) is the *additive normal development* of 1 with respect to a_1, a_2, \dots, a_n . If B be finite, and if a_1, a_2, \dots, a_n be *all* the elements of B , (ii) will be called the *complete additive normal development* of 1. The constituents in the complete additive normal development of 1 that are not 0 will be called the *minimal elements* of B , or simply the *minimals* of B .

We have at once the following basic theorem:

¹ Presented to the American Mathematical Society, June 20, 1934.

² *Transactions of the American Mathematical Society*, vol. 5 (1904), pp. 308, 309.

³ This term seems to me less simple than "minimal," and permits a less convenient name and notation for the duals of the minimal elements (See §4 below).

⁴ The reader is referred to the following more general papers on Boolean algebra, which no doubt yield, from a very different angle, many of my results: E. T. Bell, "Arithmetic of logic," *Transactions of the American Mathematical Society*, vol. 29 (1927), pp. 597-611; W. A. Hurwitz, "On Bell's arithmetic of Boolean algebra," *ibid.*, vol. 30 (1928), pp. 420-424; M. H. Stone, "On the structure of Boolean algebras" (Abstract), *Bulletin of the American Mathematical Society*, vol. 39 (1933), p. 200; M. H. Stone, "Boolean algebras and their application to topology" (Abstract), *Proceedings of the National Academy of Science*, vol. 20 (1934), pp. 197-202.

1.1. *In a finite Boolean algebra, every element a , $\neq 0$, is expressible uniquely as a sum of minimals.*

For, by (i) and (ii), every element a of the algebra is expressible uniquely as a sum of constituents in (ii), namely, the sum of those constituents in which a enters positively. Hence, if $a \neq 0$, a is expressible uniquely as the sum of those non-0 constituents in which a enters positively.

If m is a minimal in an expression for a , we shall say " m is a minimal of a ," or " m belongs to a ," or " m is contained in a ."

Theorem 1.1 enables us to construct a finite Boolean algebra from its minimals.

Theorems 1.2-1.19 below are practically all simple corollaries of 1.1 and the definition of minimals. Proofs for the more obvious of these theorems will be omitted.

1.2. *Every finite Boolean algebra has at least one minimal.*

1.3. *The order of a finite Boolean algebra is 2^k , where k is the number of minimals of the algebra.*

1.4. *If m_1, m_2, \dots, m_k are the minimals of a finite Boolean algebra, then*

$$m_1 + m_2 + \dots + m_k = 1, \quad m_i m_j = 0 \quad [i \neq j].$$

1.5. *In a finite Boolean algebra, a necessary and sufficient condition that a be a minimal is:*

$$(i) \quad a \neq 0, \quad (ii) \quad a + x \neq a \quad [x \neq 0, x \neq a].$$

For, first, let a be a minimal. Then $a \neq 0$. Also, if $a + x = a$ [$x \neq 0, x \neq a$], then x , hence $a + x$, hence a would contain minimals not belonging to a , contrary to 1.1. Hence, (i) and (ii) are necessary.

Secondly, let $a \neq 0$ and $a + x \neq a$ [$x \neq 0, x \neq a$]. Then, if a were not a minimal, we should have, by 1.1, $a = m_1 + m_2 + \dots + m_h$ [$h > 1$], where the elements m_i are minimals. Hence, there would be a minimal, say m_1 , such that $a + m_1 = a$, where $m_1 \neq 0, m_1 \neq a$, contrary to supposition. Hence, (i) and (ii) are sufficient.

1.6. *In a finite Boolean algebra, a necessary and sufficient condition that a be a minimal is:*

$$(i) \quad a \neq 0, \quad (ii) \quad x \leq a \quad [x \neq 0, x \neq a].$$

For, $x < a$ is equivalent to $a + x = a$. Hence the theorem, by 1.5.

Theorem 1.6 shows that the minimal elements are identical with Hunting-

ton's "irreducible" elements. Theorems 1.1-1.6 include all of Huntington's results concerning these elements.

1.7. Let a_1, a_2, \dots, a_k be a set of k elements of a finite Boolean algebra B such that

$$a_1 + a_2 + \dots + a_k = 1, \quad a_i \neq 0, \quad a_i a_j = 0 \quad [i \neq j];$$

if k be the number of minimals in B , then a_1, a_2, \dots, a_k are the minimals of B .

For, by 1.1, each element a_i is a minimal or a sum of minimals; and if some a_i were a sum of two or more minimals, there would be two distinct a_i 's having a minimal in common, contrary to the supposition that $a_i a_j = 0$ [$i \neq j$].

Hereafter, unless expressly stated otherwise, the algebra under consideration in any particular case will be understood to be a finite Boolean algebra, and the symbol m will always denote a minimal.

1.8. In the additive normal development of 1 with respect to any set of elements, the number of non-0 constituents cannot exceed the number of minimals.

For, each non-0 constituent is a sum of minimals, by 1.1; and no two distinct constituents, c_1 and c_2 , can have a common minimal, since $c_1 c_2 = 0$.

1.9. The sum of any set of minimals is the negative of the sum of the remaining minimals.

1.10. If m_1 and m_2 are two distinct minimals, then $m_1 m'_2 = m_1$.

For, m'_2 is of the form $m_1 + a$, by 1.9, and $m_1(m_1 + a) = m_1$.

More generally, we have:

1.11. If m_1, m_2, \dots, m_k are distinct minimals, then $m_k = m'_1 m'_2 \dots m'_{k-1} m_k$.

For, $m'_1, m'_2, \dots, m'_{k-1}$ are each of the form $m_k + a$.

1.12. In the additive normal development of 1 with respect to the minimals, the non-0 constituents are precisely the minimals.

By 1.11 and 1.8.

1.13. The minimals of $a + b$ consist of the minimals contained in a or b .

1.14. The minimals of ab consist of the minimals common to a and b .

1.15. A necessary and sufficient condition that a minimal m be a minimal of a is that $m < a$.

For, $m < a$ is equivalent to the relation $a = a + m$. Hence the theorem, by 1.1.

1. 16. If $m < a + b$, then $m < a$ or $m < b$.

By 1. 15 and 1. 13.

1. 17. $m \nless 0$.

For, if $m < 0$, then $0 = 0 + m = m$.

1. 18. Let $ab = 0$; if $m < a$ then $m \nless b$.

For, if $m < a$ and $m < b$, then $m < ab$; hence $m <$

1. 19. If $m \nless a$, then $ma = 0$.

For, if $m \nless a$, then m is not a minimal of a , by 1 by 1. 1, 1. 4.

Further properties of the minimal elements will be of the notion of "index" of an element and of other n

2. *Index of an element.* In a finite Boolean alg of an element a , denoted by $I(a)$, will be meant the belonging to a . The index of the universe element 1, minimals in B , will be called the index of B . A Boole will be denoted by B_k .

We have, for indices, theorems 2. 1-2. 5 below. (I operations between indices are, of course, arithmetic.)

2. 1. If $ab = 0$, then $I(a + b) = I(a) + I(b)$.

For, a and b cannot have a common minimal, 1 theorem, by 1. 13.

The following two theorems are corollaries of 2. 1.

2. 2. In an algebra B_k , $I(a') = k - I(a)$.

2. 3. In an algebra B_k , $I(a'b') = k - I(a + b)$.

Since if $b < a$ then $a = ab + ab' = b + ab'$, we ha

• 2. 4. If $b < a$, then $I(ab') = I(a) - I(b)$.

Since $a + b' = a + a'b = b + ab' = ab + ab' + a'b$

2. 5. $I(a + b) = I(a) + I(a'b) = I(b) + I(ab') =$

3. *Addends.* The introduction of the notion of "ciated notions will enable us to view the minimal element and will also enable us to state conveniently some addit finite Boolean algebras.

In any Boolean algebra B , if there exists an x such that $b + x = a$, then b will be called an *addend* of a , and a a *complex* of b . Two elements a, b will be said to be *relative minimals*, if 0 is the only addend common to a and b . In a finite B , the minimals of a are, of course, addends of a , and a is a complex of its minimals. If $a, \neq 0$, has no addends other than 0 and a , then, by 1. 5, a is a minimal. In a finite B , an element c will be called the *least common complex* (*L. C. C.*) of a and b , if the minimals of c are the minimals belonging either to a or to b ; an element c will be called the *highest common addend* (*H. C. A.*) of a and b , if the minimals of c are the minimals common to a and b .

3. 1. *In any Boolean algebra, a necessary and sufficient condition that b be an addend of a (or that a be a complex of b) is that $b < a$.*

For, $b < a$ is a necessary and sufficient condition that the equation $b + x = a$ have a solution.

3. 2. *In any Boolean algebra, a necessary and sufficient condition that a and b be relative minimals is that $ab = 0$.*

For, the pair of relations $x < a, x < b$ is equivalent to the single relation $x < ab$; and $ab = 0$ is a necessary and sufficient condition that $x < ab$ have the unique solution 0 .

Theorems 3. 1 and 3. 2 enable us to restate the theorems of § 1 in terms of addends and related notions. Thus, 1. 16 and 1. 18 become respectively 3. 3 and 3. 4 following.

3. 3. *If a minimal m is an addend of $a + b$, then m is an addend of a or m is an addend of b .*

3. 4. *Let a and b be relative minimals; if m is an addend of a , then m is not an addend of b .*

In view of 1. 13 and 1. 14, we have:

3. 5. $a + b = \text{L. C. C. of } a, b; ab = \text{H. C. A. of } a, b.$

The following four theorems concern the addends of an element as constituting a Boolean algebra and an abelian group.

3. 6. *In any Boolean algebra B , the addends of an element $u, \neq 0$, form a Boolean algebra A with respect to the operations $+$ and \times , the elements $0, u, ux'$ of B serving in A as the "zero," the "universe," and the "negative of x " respectively.*

For, with respect to the operation $+$, the addends of u are related to u precisely as the elements of B are related to 1.

3.7. *In any Boolean algebra B , the addends of an element u form an abelian group with respect to the operation \circ given by $x \circ y = xy' + x'y$.*

For, it can be verified, the operation \circ satisfies the following three conditions: (i) $a \circ b = b \circ a$; (ii) $(a \circ b) \circ c = a \circ (b \circ c)$; (iii) for any two addends a, b of u there is an addend x such that $a \circ x = b$.

If the algebra B in 3.6 and 3.7 be finite, these theorems become respectively Theorems 3.8 and 3.9 following.

3.8. *In an algebra B_k , the addends of an element $u, \neq 0$, of index h , form a Boolean algebra B_h with respect to the operations $+$ and \times , the elements $0, u, ux'$ of B_k serving in B_h as the "zero," the "universe," and the "negative of x " respectively.*

3.9. *In an algebra B_k , the addends of an element u , of index h , form an abelian group of order 2^h with respect to the operation \circ given by $x \circ y = xy' + x'y$.*

4. *Dual considerations. The maximals:* The foregoing considerations have, of course, their dual counterparts. Thus, corresponding to "minimal," "addend," "complex," "relative minimals," "least common complex," "highest common addend," we have, respectively, "maximal," "factor," "multiple," "relative maximals," "least common multiple," "highest common factor." The definitions of the notions in the latter set follow.

A *maximal* element μ of a finite Boolean algebra B is a non-1 constituent in the *complete* multiplicative normal development of 0. In any Boolean algebra, if there exists an x such that $bx = a$, then b is a *factor* of a , and a is a *multiple* of b ; two elements a, b are *relative maximals* if 1 is the only factor common to a and b . In a finite Boolean algebra, an element c is the *least common multiple* (L.C.M.) of a and b , if the maximals of c are the maximals belonging either to a or to b ; an element c is the *highest common factor* (H.C.F.) of a and b , if the maximals of c are the maximals common to a and b .

As samples of the duals of the foregoing propositions, I give propositions 4.1-4.5 below. These are the duals of 1.1, 1.4, 1.6, 1.10, 3.1 respectively.

4.1. *In a finite Boolean algebra, every element $a, \neq 1$, is expressible uniquely as a product of maximals.*

4.2. *If $\mu_1, \mu_2, \dots, \mu_k$ are the maximals of a finite Boolean algebra, then*

$$\mu_1 \mu_2 \dots \mu_k = 0, \quad \mu_i + \mu_j = 1 \quad [i \neq j].$$

4.3. In a finite Boolean algebra, a necessary and sufficient condition that a be a maximal is:

$$(i) \quad a \neq 1, \quad (ii) \quad x \not\geq a \quad [x \neq 1, x \neq a].$$

4.4. If μ_1 and μ_2 are two distinct maximals, then $\mu_1 + \mu'_2 = \mu_1$.

4.5. In any Boolean algebra, a necessary and sufficient condition that b be a factor of a (or that a be a multiple of b) is that $b > a$.

In view of 3.1, Theorem 4.5 tells us:

4.6. In any Boolean algebra, a necessary and sufficient condition that a be a multiple of b is that a be an addend of b .

5. *Relation between the minimals and the maximals.* As is to be expected, there is a very close relation between the minimal and the maximal elements of a finite Boolean algebra. This relation is given by the following theorem.

5.1. In a finite Boolean algebra, if m is a minimal then m' is a maximal.

For, if m is a non-0 constituent in the complete additive normal development of 1, then m' is a non-1 constituent in the complete multiplicative normal development of 0.

As a consequence of 5.1, in view of 1.10, we have:

5.2. In a finite Boolean algebra, let m be a minimal and μ a maximal; then $m\mu = 0$, or else $m\mu = m$.

Dually, we have:

5.3. In a finite Boolean algebra, let μ be a maximal and m a minimal; then $\mu + m = 1$, or else $\mu + m = \mu$.

6. *Existence and construction of sub-algebras.* I shall now apply the properties of minimals to answer some questions concerning sub-algebras of a finite Boolean algebra.

We know that every Boolean algebra B has "embedded" in it as a sub-algebra the algebra B_1 of index 1, consisting of the elements 0, 1. Are there sub-algebras B_h [$h > 1$] embedded in an algebra B_k [$k > h$]? If sub-algebras B_h exist, how construct these algebras? Again, given an algebra B_h embedded in a B_k [$k > h + 1$], does there exist an algebra B_{h+1} embedded in B_k and embedding B_h ? The answers to these questions will be found in the Theorems 6.1-6.3 following.

6.1. Let m_1, m_2, \dots, m_k be the minimals of an algebra B_k of $k-1$ elements

$$m_1, m_2, \dots, m_{k-2}, m_{k-1} + m_k$$

constitute a set of $k-1$ minimals for a sub-algebra B_k

The theorem is true by virtue of 1.7.

As a corollary of 6.1, we have:

6.2. Every algebra B_k has embedded in it sub-algebras B_h [$1 \leq h \leq k$].

Theorems 6.1 and 6.2 enable us to construct systems of algebras of a given finite Boolean algebra.

6.3. Let m_1, m_2, \dots, m_h be the minimals of an algebra B_k [$k > h$]; there is in B_k an element minimal, say m_h , the $h+1$ elements

$$m_1, m_2, \dots, m_{h-1}, xm_h, x'm_h$$

constitute a set of minimals for an algebra B_{h+1} embedded

in B_k . For, there exists in B_k some minimal of B_h , say m_h , element $x, \neq 0, \neq m_h$, such that $x < m_h$; otherwise, m_h is in B_k an element x such that $xm_h = x$ [$x \neq 0, x \neq m_h$]. In B_k an element x such that $xm_h \neq 0$ [$x \neq 0, x \neq m_h$]. have $x'm_h \neq 0$, i. e. $m_h \leq x$; otherwise, $x = m_h$. Since x and the elements $m_1, m_2, \dots, m_{h-1}, xm_h, x'm_h$ form, by 1.7, an algebra B_{h+1} ; and this algebra embeds in B_k .

The reader can readily supply the duals of the above.

7. *A problem in dichotomy.* In a Boolean algebra constituents in the additive normal development of 1 will, in general, vanish. The following problem then arises.

PROBLEM. In an algebra B_k , to find a set of h elements which none of the additive constituents of 1 vanish.

The properties of the minimal elements established in the previous sections enable us to solve this problem readily. We observe that our problem has a solution, we must have, by 1.8, $k \geq 2^h$. Then, by 6.2, there exists in B_k a set of 2^h elements

which can serve as a set of minimals for a sub-algebra B_2^h embedded in B_k . The desired h elements are the elements a_i determined by the following $2^h - 1$ equations:

$$a_1 a_2 \cdots a_h = m_1, a_1 a_2 \cdots a'_h = m_2, \cdots, a'_1 a'_2 \cdots a_h = m_{2^h-1},$$

where the a -products are all different, and where $a'_1 a'_2 \cdots a'_h$ does not appear.

Thus, consider a Boolean algebra B_8 . Let the minimals of B_8 be m_1, m_2, \cdots, m_8 . A set of three elements with respect to which none of the additive constituents of 1 vanish are a_1, a_2, a_3 given by the following $2^3 - 1$ equations:

$$\begin{cases} a_1 a_2 a_3 = m_1, & a_1 a_2 a'_3 = m_2, & a_1 a'_2 a_3 = m_3, \\ a'_1 a_2 a_3 = m_4, & a_1 a'_2 a'_3 = m_5, & a'_1 a_2 a'_3 = m_6, \\ a'_1 a'_2 a_3 = m_7. \end{cases}$$

Indeed, noting that the sum of the products involving a_i affirmatively is a_i , we have:

$$\begin{cases} a_1 = m_1 + m_2 + m_3 + m_5, & a_2 = m_1 + m_2 + m_4 + m_6, \\ a_3 = m_1 + m_3 + m_4 + m_7. \end{cases}$$

The reader can supply the dual of the above problem and its solution.

8. *Arithmetic representations of finite Boolean algebras.* The properties of the minimal elements and related concepts deliver into our hands very simple arithmetic representations of finite Boolean algebras. These representations are given by the following considerations.

Let N_k be the class of 2^k numbers consisting of the number 1 and the products of h distinct primes taken from a set of k distinct primes p_1, p_2, \cdots, p_k [$1 \leq h \leq k$]. Further, let $a \square b$ be the least common multiple of a and b , and $a \Delta b$ the greatest common divisor of a and b . Then, no number of N_k contains a prime to a power higher than 1. Hence, the primes of N_k play with regard to \square and Δ the precise rôle that the minimals of B_k play with regard to Boolean addition, \oplus , and Boolean multiplication, \odot , respectively. Hence, the arithmetic system (N_k, \square, Δ) is simply isomorphic with the Boolean system (B_k, \oplus, \odot) , \square corresponding to \oplus , and Δ to \odot .⁵

Dually, the primes of N_k play with regard to \square and Δ the precise rôle that the maximals of B_k play with regard to Boolean \odot and \oplus respectively.

⁵ The Boolean "zero," "universe," and "negative of a " correspond respectively to the N_k -elements 1, $v = p_1 \square p_2 \square \cdots \square p_k$, $a_1 = v \div a$.

Hence, the arithmetic system (N_k, \square, Δ) is simply isomorphic with the Boolean system (B_k, \odot, \oplus) , \square corresponding to \odot and Δ to \oplus .⁶

It is to be observed that the representations of finite Boolean algebras given here enable us to "number" conveniently the elements of a finite Boolean algebra. Thus, to number the elements of an algebra B_3 , we take for the three minimals in B_3 any three primes, for simplicity, the first three primes: 2, 3, 5. The numbered elements of B_3 are, then:

$$\begin{aligned} 1, 2, 3, 5, 6 (= 2 \square 3), 10 (= 2 \square 5), 15 (= 3 \square 5), \\ 30 (= 2 \square 3 \square 5 = v).^7 \end{aligned}$$

9. *New proof of the order theorem.* I shall close my discussion of finite Boolean algebras with a new proof of the theorem that *the order of a finite Boolean algebra is of the form 2^k [$k > 0$]*. This proof has interest in that no use is made in it of the minimal (or the maximal) elements. The proof follows.

Let A be a finite Boolean algebra of order n . Consider any group operation in A , say the operation \circ given by $a \circ b = ab' + a'b$. With respect to the elements of A form an abelian group. Suppose, now, that A has a sub-group B of order 2^h [$h \geq 1$]. Let b_1, b_2, \dots, b_{2^h} be the elements of B . Then, for any element c of A outside B , the 2^{h+1} elements

$$b_1, b_2, \dots, b_{2^h}, b_1 \circ c, b_2 \circ c, \dots, b_{2^h} \circ c$$

form in A an abelian group of order 2^{h+1} . Hence, since A is finite, the number n of its elements must be of the form 2^k [$k > 0$]. That is, we find that if A has any sub-group of order 2^h [$h \geq 1$], then the order of A is of the form 2^k [$k > 0$]. But A has the sub-group of order 2, consisting of the elements 0, 1. Hence the theorem.

THE UNIVERSITY OF CALIFORNIA.

⁶ The Boolean "universe," "zero," and "negative" correspond respectively to the N_k -elements 1, v , a_1 .

⁷ Compare Sheffer's representations of finite Boolean algebras, reproduced by Huntington, *Transactions of the American Mathematical Society*, vol. 35 (1933), p. 278.

LINEAR METRIC LEBESGUE SPACES.

By ANDREW C. BERRY.

The present paper is intended to serve as an introduction to a study of the elements of linear functional analysis in certain function-spaces whose metrics are given explicitly in terms of Lebesgue integrals. Spaces $L_{\phi, \psi}$, generalizations of spaces L_p of functions whose p -th powers are integrable, are introduced. Each such is here shown to be a complete linear metric space. Simple necessary and sufficient conditions for separability are found.

In the classical theory, spaces L_p , L_q , $L_{pq/(p+q)}$ are connected by an inequality due to Hölder. Using this fact, M. Riesz investigated convexity relations involving the individual members of families of such spaces. The indicated generalizations will be studied in a second paper.

Finally, by means of these new results, a third paper will extend the elementary theory of linear functionals from the classical symmetric case to the more nearly general case involving pairs of distinct spaces $L_{\phi, \psi}$.

I. THE GENERAL HYPOTHESES AND THE FUNDAMENTAL DEFINITION.

1.1. *Introduction.* It will be assumed that the reader is familiar with those laws of vector-algebra (dealing with addition of two elements and multiplication of an element by a complex number) which serve as postulates for a linear abstract space.¹ It will be assumed, further, that the reader is familiar with the details of the process of associating with each given Lebesgue-measurable function $f(x)$ every measurable function coinciding almost everywhere with $f(x)$, terming the totality thus formed a function-element, and extending algebraic operations and Lebesgue integration to these elements. Wherefore, the reader will allow us to avoid these precise constructions and to adopt in their stead the convenient, albeit paradoxical usage which considers each measurable function as an element and yet refuses to distinguish between elements $f(x)$ and $g(x)$ which, as functions, differ on a set of measure zero.

We shall investigate certain function-spaces $L_{\phi, \psi}$ each of which is determined by a metric-generating pair

¹ For complete details consult: N. Wiener, *Bulletin de la Société Mathématique de France*, t. 49 (1921), pp. 119-134; S. Banach, *Fundamenta Mathematicae*, t. 3 (1922), pp. 133-181; M. Fréchet, *Les espaces abstraits*, Paris, Gauthier-Villars (1928), particularly pp. 125-126; S. Banach, *Théorie des opérations linéaires*, Warsaw, Mathematical Seminar of the University of Warsaw (1932).

$$\phi(u), \quad \psi(u)$$

of functions which, it will be understood throughout this paper, satisfy the following three requirements.

(1.11) $\phi(u)$ shall be defined for $0 \leq u \leq \infty$ as a non-negative, non-decreasing function of u , with $+\infty$ as a permissible value:

$$0 \leq \phi(u) \leq \phi(v) \leq \infty, \quad (0 \leq u < v \leq \infty).$$

Furthermore, it shall be required that

$$\phi(0) = 0, \quad \phi(1) = 1, \quad \phi(\infty) = \lim_{u \rightarrow \infty} \phi(u).$$

(1.12) $\psi(u)$ shall be defined for $0 < u < \infty$ as a non-negative, non-decreasing function of u , with $+\infty$ as a permissible value:

$$0 \leq \psi(u) \leq \psi(v) \leq \infty, \quad (0 < u < v < \infty).$$

Furthermore, it shall be required that

$$0 < \psi(1) < \infty.$$

(1.13) The inequality

$$\psi(sw)\psi(tw)\phi(su + tv) \leq \psi(w)\{s\psi(tw)\phi(u) + t\psi(sw)\phi(v)\}$$

shall be satisfied whenever all the functional values involved are finite and

$$0 < s, \quad 0 < t, \quad s + t = 1, \quad 0 \leq u < \infty, \quad 0 \leq v < \infty, \quad 0 < w < \infty.$$

Definition 1.14. $L_{\phi, \psi}$ will denote the totality of complex-valued Lebesgue-measurable functions, $f(x)$, of the single real variable x , $-\infty < x < \infty$, for each of which there exists, as a finite number,

$$l_{\phi, \psi}(f) = \inf \epsilon > 0 \text{ such that } \int \phi \left(\frac{|f(x)|}{\epsilon} \right) dx \leq \psi(\epsilon),$$

the integral being a Lebesgue integral² extended over $(-\infty, \infty)$. If, for a given ϵ , it is true that $\psi(\epsilon) = \infty$, then the inequality of this definition is to be considered satisfied for this ϵ .

1.2. Discussion of the general hypotheses.

For convenience, we introduce the following standard notation.

² It is known that a monotone function of a measurable function of a real variable is itself a measurable function of the variable. See, for example, Carathéodory, *Vorlesungen über reelle Funktionen*, 2nd ed. (1927), B. G. Teubner, Berlin, §§ 348, 351.

$$\left\{ \begin{array}{l} \phi(u+) = \lim_{\substack{\epsilon \rightarrow 0 \\ (\epsilon > 0)}} \phi(u + \epsilon), \quad \psi(u+) = \lim_{\substack{\epsilon \rightarrow 0 \\ (\epsilon > 0)}} \psi(u + \epsilon), \quad (0 \leq u < \infty). \\ \phi(u-) = \lim_{\substack{\epsilon \rightarrow 0 \\ (0 < \epsilon < u)}} \phi(u - \epsilon), \quad \psi(u-) = \lim_{\substack{\epsilon \rightarrow 0 \\ (0 < \epsilon < u)}} \psi(u - \epsilon), \quad (0 < u < \infty). \end{array} \right.$$

THEOREM 1.21. If $\phi(0+) > 0$, then $\psi(0+) = 0$.

Proof. In (1.13) we set $u = 0$, $v = 1$, $w = 1$, and so find

$$\psi(s)\psi(t)\phi(t) \leq \psi(1)[t\psi(s)] \leq [\psi(1)]^2 t.$$

Noting that $\psi(t) \leq \psi(s)$ when $0 < t \leq \frac{1}{2}$, we conclude that

$$[\psi(t)]^2 \leq \frac{[\psi(1)]^2}{\phi(0+)} t, \quad (0 < t \leq \tfrac{1}{2}).$$

This result establishes our theorem.

We shall now show that the two requirements: $\phi(1) = 1$, $0 < \psi(1) < \infty$, introduce no essential loss of generality. We note, in passing, that the condition $\phi(0) = 0$ makes certain that the null element (i. e., "the" function which is zero almost everywhere) will have the length zero. That there may be no second element of length zero it is necessary to require both that $\phi(u)$ be not identically zero in its domain of definition and that $\psi(u)$ possess at least one finite value. Thus, there must exist u_1 and v_1 such that

$$\begin{array}{ll} 0 < u_1 < \infty, & 0 < \phi(u_1); \\ 0 < v_1 < \infty, & \psi(v_1) < \infty. \end{array}$$

Next, it is desirable that $L_{\phi, \psi}$ fail to be a null space (i. e., a space consisting solely of the null element). One necessary condition for such failure (since it has just been shown that $\phi(u_1) > 0$) is that $\psi(u)$ have at least one positive value. Thus, there must exist v_2 such that

$$0 < v_2 < \infty, \quad 0 < \psi(v_2).$$

A second condition, necessary in the case in which $\psi(u)$ is everywhere finite, is: $\phi(0+) < \infty$.

Another degenerate type of space $L_{\phi, \psi}$, also sufficiently trivial to warrant rejection, is that which contains as elements all measurable functions $f(x)$ and assigns to all its elements, other than the null element, one constant length l , $0 < l < \infty$. The conditions: $\phi(0+) = \infty$, $\psi(u) = \infty$ for some u , are separately necessary and collectively sufficient that $L_{\phi, \psi}$ be such a space (with l defined by the relations $\psi(l-) < \infty$, $\psi(l+) = \infty$).

In view of the foregoing remarks we finally must require $\phi(0+)$ to be finite. Thus, there must exist u_2 such that

$$0 < u_2 < \infty, \quad \phi(u_2) < \infty.$$

On combining our several results we readily demonstrate that there must exist u_0 and v_0 such that

$$\begin{aligned} 0 < u_0 < \infty, & \quad 0 < \phi(u_0 +), & \quad \phi(u_0 -) < \infty; \\ 0 < v_0 < \infty, & \quad 0 < \psi(v_0 +), & \quad \psi(v_0 -) < \infty. \end{aligned}$$

A brief consideration of the fundamental definition shows that we may alter, if necessary, the values $\phi(u_0)$ and $\psi(v_0)$ so that

$$0 < \phi(u_0) < \infty, \quad 0 < \psi(v_0) < \infty,$$

without altering the remaining values of these functions, and so without altering the space $L_{\phi, \psi}$.

Now, introduce the functions

$$\phi_1(u) = \frac{\phi(u_0 u)}{\phi(u_0)}, \quad \psi_1(u) = \frac{\psi(v_0 u)}{\psi(v_0)}.$$

We see, firstly, that

$$\phi_1(1) = 1, \quad 0 < \psi_1(1) < \infty,$$

and, secondly, that

$$l_{\phi, \psi}(u_0 v_0 f) = v_0 l_{\phi_1, \psi_1}(f).$$

The two spaces $L_{\phi, \psi}$ and L_{ϕ_1, ψ_1} are essentially equivalent, the one being merely a homogeneous enlargement of the other (partially from the point of view of the metric assigned to a given element and partially from the point of view of the elements themselves). Thus, the aim of the present discussion has been attained.

1.3. Examples of spaces $L_{\phi, \psi}$.

THEOREM 1.31. *If $\phi(u)$ satisfies (1.11) and if $\psi(u) = u$, ($0 < u < \infty$), then the general hypotheses are fulfilled for the pair $\phi(u)$, $\psi(u)$.*

Proof. The function $\psi(u) = u$ satisfies (1.12) and reduces (1.13) to

$$\phi(su + tv) \leq \phi(u) + \phi(v),$$

which automatically is satisfied since $(su + tv)$ is intermediate in value between u and v and since $\phi(u)$ is non-negative and monotone.

THEOREM 1.32. *If $\phi(u)$ satisfies (1.11) and is a "convex" function of u (i. e., satisfies the inequality: $\phi(su + tv) \leq s\phi(u) + t\phi(v)$, whenever $0 < s$, $0 < t$, $s + t = 1$), and if $\psi(u) = \text{constant}$ ($0 < \text{const} < \infty$) for $0 < u < \infty$, then the pair $\phi(u)$, $\psi(u)$ is satisfactory.*

Proof. Immediate.

Example 1.33. By 1.32, the pair

$$\begin{cases} \phi(u) = u^p & (\text{for some } p \text{ such that } 1 \leq p < \infty), \\ \psi(u) = 1, & (0 < u < \infty), \end{cases} \quad (0 \leq u \leq \infty),$$

is known to satisfy the general hypotheses. A simple calculation yields:

$$l_{\phi, \psi}(f) = \left\{ \int |f(x)|^p dx \right\}^{1/p}$$

Thus, the corresponding $L_{\phi, \psi}$ is the familiar space of measurable functions whose p -th powers are integrable.³ This is known, in particular, to be a separable, complete, linear metric space with homogeneous metric.

Example 1.34. Consider the pair

$$\begin{cases} \phi(u) = u^q & (\text{for some } q \text{ such that } 0 < q < 1), \\ \psi(u) = u^{1-q}, & (0 < u < \infty). \end{cases} \quad (0 \leq u \leq \infty),$$

Here, (1.13) becomes:

$$(su + tv)^q \leq (su)^q + (tv)^q.$$

The reader will establish the validity of this inequality. The corresponding space later will be seen to be separable, complete, and linear metric. However, the metric

$$l_{\phi, \psi}(f) = \int |f(x)|^q dx$$

is not homogeneous.

Example 1.35. For the, by 1.32, satisfactory pair

$$\begin{cases} \phi(u) = \lim_{p \rightarrow \infty} u^p = \begin{cases} 0, & (0 \leq u < 1), \\ 1, & (u = 1), \\ \infty, & (1 < u \leq \infty), \end{cases} \\ \psi(u) = 1, & (0 < u < \infty), \end{cases}$$

we find

$$\begin{aligned} l_{\phi, \psi}(f) &= \text{upper measurable bound of } |f(x)| \\ &= \inf \epsilon > 0 \text{ such that } \text{meas} \{ |f(x)| > \epsilon \} = 0. \end{aligned}$$

This metric is homogeneous and the corresponding $L_{\phi, \psi}$ is the familiar non-separable, complete, linear metric space of measurably bounded functions.

³ F. Riesz, "Über Systeme integrierbarer Funktionen," *Mathematische Annalen*, Bd. 69 (1910), pp. 449-497; Banach, *Théorie des opérations linéaires*, cited above.

Example 1.36. The, by 1.31, satisfactory pair

$$\begin{cases} \phi(u) = \lim_{\substack{q \rightarrow 0 \\ (q > 0)}} u^q = \begin{cases} 0, & (u = 0), \\ 1, & (0 < u < \infty), \end{cases} & \phi(\infty) = 1, \\ \psi(u) = u, & (0 < u < \infty), \end{cases}$$

yields the non-homogeneous metric

$$l_{\phi, \psi}(f) = \text{meas} \{ |f(x)| > 0 \}.$$

Example 1.37. The pair

$$\begin{cases} \phi(u) = \begin{cases} 0, & (0 \leq u < 1), \\ 1, & (1 \leq u \leq \infty), \end{cases} \\ \psi(u) = u, & (0 < u < \infty), \end{cases}$$

determines the metric

$$l_{\phi, \psi}(f) = \inf \epsilon > 0 \text{ such that } \text{meas} \{ |f(x)| > \epsilon \} \leq \epsilon.$$

The corresponding space $L_{\phi, \psi}$ has been discussed by us elsewhere.*

II. EACH $L_{\phi, \psi}$ IS A COMPLETE LINEAR METRIC SPACE.

If $f(x)$ is an element of a space $L_{\phi, \psi}$ having the property S (which property will be defined in 3.11), then $f(x)$ must be finite almost everywhere. For a general $L_{\phi, \psi}$, however, a linear combination of elements, $af(x) + bg(x)$, may involve the indeterminate forms: $\infty - \infty$, $0 \cdot \infty$. On the set of indeterminate values the linear combination may be defined according to our pleasure by equating it to any measurable function. The modifications thus necessitated in the linear algebra are well known.

2.1. Preliminary estimates.

THEOREM 2.11. *If $f(x)$ is an element of a given $L_{\phi, \psi}$ and if $\epsilon > l_{\phi, \psi}(f)$, then the inequality of 1.14 will be satisfied for this $f(x)$ and this ϵ .*

Proof. By 1.14, there exists ϵ_1 , $l_{\phi, \psi}(f) \leq \epsilon_1 < \epsilon$, such that

$$\int \phi \left(\frac{|f(x)|}{\epsilon_1} \right) dx \leq \psi(\epsilon_1).$$

* Berry, *Proceedings of the National Acad. of Sciences*, vol. 17 (1931), pp. 456-459. I take this opportunity to make a correction. In this note I stated that the metric

$$l(f) = \inf_{(\epsilon > 0)} (\epsilon + \text{meas} \{ |f(x)| > \epsilon \}),$$

introduced by M. Fréchet ("Sur divers modes de convergence d'une suite de fonctions d'une variable," *Bulletin of the Calcutta Mathematical Society*, vol. 11 (1921), pp. 187-206; *Les espaces abstraits*, cited above, p. 91) was said by him not to satisfy the triangle inequality 2.22. Secondly, I impliedly agreed with this supposedly quoted opinion.

By (1. 11) and (1. 12) the desired inequality,

$$\int \phi \left(\frac{|f(x)|}{\epsilon} \right) dx \leq \psi(\epsilon),$$

follows.

THEOREM 2. 12. *If $f(x)$ is an element of a given space $L_{\phi, \psi}$, if $g(x)$ is measurable, and if $|g(x)| \leq |f(x)|$, almost everywhere, then $g(x)$ is an element of $L_{\phi, \psi}$, and $l_{\phi, \psi}(g) \leq l_{\phi, \psi}(f)$.*

Proof. By (1. 11), it is seen that the inequality of 1. 14, if satisfied for the given $f(x)$ and for a given $\epsilon > 0$, also will be satisfied for $g(x)$ and the same ϵ . This fact establishes the theorem.

THEOREM 2. 13. *If $f(x)$ is an element of a given $L_{\phi, \psi}$, if $0 \leq l_{\phi, \psi}(f) < \frac{1}{4}$, and if $u_0, 0 \leq u_0 < \infty$ is such that $0 < \phi(u_0 +)$, then*

$$\text{meas} \{ |f(x)| > u_0 \sqrt{l_{\phi, \psi}(f)} \} \leq \frac{\psi(1)}{\phi(u_0 +)} \sqrt{l_{\phi, \psi}(f)}.$$

Proof. Let ϵ and η be chosen so that

$$0 \leq l_{\phi, \psi}(f) < \epsilon < \frac{1}{4}, \quad 0 < \eta < \infty,$$

and let these values be kept fixed throughout the discussion of the following three cases.

Case A. $\psi(\epsilon) = 0$.

Here, by 2. 11,

$$\phi \left(\frac{|f(x)|}{\epsilon} \right) = 0, \text{ almost everywhere, } (-\infty < x < \infty).$$

Since

$$\phi \left(\frac{u_0 + \eta}{\epsilon^{1/2}} \right) \geq \phi(u_0 +) > 0,$$

it follows that

$$\text{meas} \{ |f(x)| > (u_0 + \eta)\epsilon^{1/2} \} = 0.$$

Case B. $\phi \left(\frac{u_0 + \eta}{\epsilon^{1/2}} \right) = \infty$.

Here, since 2. 11 yields:

$$\int \phi \left(\frac{|f(x)|}{\epsilon} \right) dx \leq \psi(\epsilon) \leq \psi(1) < \infty,$$

we see that

$$\text{meas} \{ |f(x)| > (u_0 + \eta)\epsilon^{1/2} \} = 0.$$

Case C. $\psi(\epsilon) > 0$ and $\phi\left(\frac{u_0 + \eta}{\epsilon^{1/2}}\right) < \infty$.

Here we note that

$$0 < \phi(u_0 + \eta) \leq \phi(u_0 + \eta) \leq \phi\left(\frac{u_0 + \eta}{\epsilon^{1/2}}\right)$$

$$0 < \psi(\epsilon) \leq \psi(\epsilon^{1/2}) \leq \psi(1 - \epsilon^{1/2}) \leq \psi(1).$$

Thus, if in (1.13) we set $u = 0$, $v = \left(\frac{u_0 + \eta}{\epsilon^{1/2}}\right)$, $w =$
we find that

$$1 \leq \frac{\psi(1)}{\phi(u_0 + \eta)} \epsilon^{1/2} \frac{\phi\left(\frac{u_0 + \eta}{\epsilon^{1/2}}\right)}{\psi(\epsilon)}.$$

Whence,

$$\begin{aligned} \text{meas } \{|f(x)| > (u_0 + \eta)\epsilon^{1/2}\} \\ &\leq \frac{\psi(1)}{\phi(u_0 + \eta)} \epsilon^{1/2} \frac{\phi\left(\frac{u_0 + \eta}{\epsilon^{1/2}}\right) \text{meas } \{|f(x)| > (u_0 + \eta)\epsilon^{1/2}\}}{\psi(\epsilon)} \\ &\leq \frac{\psi(1)}{\phi(u_0 + \eta)} \epsilon^{1/2} \frac{\int \phi\left(\frac{|f(x)|}{\epsilon}\right) dx}{\psi(\epsilon)} \leq \frac{\psi(1)}{\phi(u_0 + \eta)} \end{aligned}$$

In all cases we have found that the inequality

$$\text{meas } \{|f(x)| > (u_0 + \eta)\epsilon^{1/2}\} \leq \frac{\psi(1)}{\phi(u_0 + \eta)}$$

holds for each sufficiently small $\epsilon > l_{\phi, \psi}(f)$ and for each only such values of ϵ and η , but requiring

$$\epsilon \rightarrow l_{\phi, \psi}(f), \quad \eta \rightarrow 0,$$

we see that the set $\{|f(x)| > (u_0 + \eta)\epsilon^{1/2}\}$ expands
 $\{|f(x)| > u_0 \sqrt{l_{\phi, \psi}(f)}\}$. Under such circumstances we
 $\Rightarrow \lim \text{meas}$. Therefore,

$$\text{meas } \{|f(x)| > u_0 \sqrt{l_{\phi, \psi}(f)}\} \leq \frac{\psi(1)}{\phi(u_0 + \eta)} \sqrt{l_{\phi, \psi}(f)}$$

2.2. Proof that each $L_{\phi, \psi}$ is a linear metric space.

THEOREM 2.21. A necessary and sufficient condition is that $f(x) = 0$ almost everywhere, i. e., that f be the n

Proof of sufficiency. Let $f(x) = 0$ almost everywhere. Then for each $\epsilon > 0$,

$$\int \phi\left(\frac{|f(x)|}{\epsilon}\right) dx = 0 \leq \psi(\epsilon).$$

By definition 1.14, $l_{\phi,\psi}(f) = 0$.

Proof of necessity. Let $l_{\phi,\psi}(f) = 0$. Since $\phi(1+) > 0$, it follows from 2.13 that

$$\text{meas}\{|f(x)| > 0\} = 0,$$

which means precisely that $f(x) = 0$, almost everywhere.

THEOREM 2.22. *If $f(x)$ and $g(x)$ are both elements of a given $L_{\phi,\psi}$, then the "triangle inequality,"*

$$l_{\phi,\psi}(f+g) \leq l_{\phi,\psi}(f) + l_{\phi,\psi}(g),$$

is valid and implies that $f(x) + g(x)$ also is an element of $L_{\phi,\psi}$.

Proof. Let ϵ and η be chosen so that

$$l_{\phi,\psi}(f) < \epsilon < \infty, \quad l_{\phi,\psi}(g) < \eta < \infty,$$

and let these values be kept fixed during the discussion of the three following cases.

Case A. $\psi(\epsilon + \eta) = \infty$.

Here we are to regard the inequality

$$\int \phi\left(\frac{|f(x) + g(x)|}{\epsilon + \eta}\right) dx \leq \psi(\epsilon + \eta)$$

as satisfied.

Case B. At least one of the values $\psi(\epsilon)$, $\psi(\eta)$ is zero.

Without loss of generality we may assume that $\psi(\epsilon) = 0$. This, in virtue of 2.11, implies that

$$\phi\left(\frac{|f(x)|}{\epsilon}\right) = 0, \text{ almost everywhere, } (-\infty < x < \infty).$$

Since $\phi(u)$ is non-decreasing, we conclude that

$$\begin{aligned} & \phi\left(\frac{|f(x) + g(x)|}{\epsilon + \eta}\right) \leq \phi\left(\frac{|f(x)| + |g(x)|}{\epsilon + \eta}\right) \\ & \leq \max\left[\phi\left(\frac{|f(x)|}{\epsilon}\right), \phi\left(\frac{|g(x)|}{\eta}\right)\right] = \phi\left(\frac{|g(x)|}{\eta}\right), \text{ almost everywhere.} \end{aligned}$$

Thus,

$$\int \phi\left(\frac{|f(x) + g(x)|}{\epsilon + \eta}\right) dx \leq \int \phi\left(\frac{|g(x)|}{\eta}\right) dx$$

Case C. $0 < \psi(\epsilon)$, $0 < \psi(\eta)$ and $\psi(\epsilon + \eta) <$

If, in (1.13), we set

$$u = \frac{|f(x)|}{\epsilon}, \quad v = \frac{|g(x)|}{\eta}, \quad w = \epsilon + \eta, \quad s =$$

we find that

$$\begin{aligned} & \phi\left(\frac{|f(x) + g(x)|}{\epsilon + \eta}\right) \leq \phi\left(\frac{|f(x)|}{\epsilon} + \frac{|g(x)|}{\eta}\right) \\ & \leq \psi(\epsilon + \eta) \left\{ \frac{\epsilon}{\epsilon + \eta} \frac{\phi\left(\frac{|f(x)|}{\epsilon}\right)}{\psi(\epsilon)} + \frac{\eta}{\epsilon + \eta} \frac{\phi\left(\frac{|g(x)|}{\eta}\right)}{\psi(\eta)} \right\} \end{aligned}$$

Since 2.11 yields:

$$\int \phi\left(\frac{|f(x)|}{\epsilon}\right) dx \leq \psi(\epsilon), \quad \text{and} \quad \int \phi\left(\frac{|g(x)|}{\eta}\right) dx \leq \psi(\eta)$$

it follows that

$$\int \phi\left(\frac{|f(x) + g(x)|}{\epsilon + \eta}\right) dx \leq \psi(\epsilon + \eta)$$

Collecting results, we find that this last inequality holds for all $\epsilon > l_{\phi, \psi}(f)$ and all $\eta > l_{\phi, \psi}(g)$. We conclude, therefore, that

$$l_{\phi, \psi}(f + g) \leq \epsilon + \eta$$

for all $\epsilon > l_{\phi, \psi}(f)$ and all $\eta > l_{\phi, \psi}(g)$. This is possible if and only if

$$l_{\phi, \psi}(f + g) \leq l_{\phi, \psi}(f) + l_{\phi, \psi}(g)$$

THEOREM 2.23. *If $f(x)$ is in $L_{\phi, \psi}$ and if a is a constant, then $af(x)$ is in $L_{\phi, \psi}$, and*

$$\begin{cases} l_{\phi, \psi}(af) \leq |a| l_{\phi, \psi}(f), & \text{when } |a| \geq 1 \\ l_{\phi, \psi}(af) = l_{\phi, \psi}(f), & \text{when } |a| = 1 \\ l_{\phi, \psi}(af) \leq l_{\phi, \psi}(f), & \text{when } |a| < 1 \end{cases}$$

In the special case in which $\psi(u) = \text{constant}$, ($0 < \text{constant}$), the inequality can be strengthened to the equality $l_{\phi, \psi}(af) = |a| l_{\phi, \psi}(f)$, i. e., the metric is homogeneous in this case.

Proof. We need only consider the case: $|a| > 0$.

When $\psi(u) = \text{constant}$, the inequalities

$$\int \phi \frac{|f(x)|}{\epsilon} dx \leq \psi(\epsilon), \quad \int \phi \frac{|af(x)|}{|a|\epsilon} dx \leq \psi(|a|\epsilon),$$

have their left members equal and their right members equal, and so are equivalent in the sense that whenever one is valid so also is the other. By 1. 14, then, $l_{\phi, \psi}(af) = |a| l_{\phi, \psi}(f)$, for all a .

In general we see that these same two inequalities are equivalent only when $|a| = 1$, but that the validity of the first implies that of the second whenever $|a| > 1$. These facts establish, respectively, the second and first relations of the theorem. The third relation is an immediate consequence of 2. 12.

2. 3. An existence theorem.

THEOREM 2. 31. *If $g_1(x), g_2(x), \dots$, are elements of a given $L_{\phi, \psi}$; if (for a given constant, T , independent of n) it is true that $g_n(x) = 0$ whenever $|x| > T$, ($n = 1, 2, \dots$); if*

$$|g_{n+1}(x) - g_n(x)| \leq 2^{-n-1}, \text{ for all } x \text{ and all } n;$$

and if for each p , ($p = 1, 2, \dots$),

$$l_{\phi, \psi}(g_m - g_n) \leq 2^{-2p-2}, \text{ when } m \geq p \text{ and } n \geq p;$$

then there exists an element $g(x)$ in $L_{\phi, \psi}$ such that

$$\lim_{n \rightarrow \infty} l_{\phi, \psi}(g - g_n) = 0.$$

Proof. We note first that, since

$$g_n(x) = g_1(x) + \{g_2(x) - g_1(x)\} + \dots + \{g_n(x) - g_{n-1}(x)\},$$

there exists

$$g(x) = \lim_{n \rightarrow \infty} g_n(x) = g_1(x) + \{g_2(x) - g_1(x)\} + \dots,$$

defined for all x as the sum of a uniformly and absolutely convergent series. In particular, we see that

$$|g(x) - g_n(x)| \leq |g_{n+1}(x) - g_n(x)| + |g_{n+2}(x) - g_{n+1}(x)| + \dots \leq 2^{-n},$$

for all x and all n .

Now let $\epsilon > 0$ be given.

Case A. $\psi(\epsilon) = 0$.

Here, determine n_ϵ so that

$$2^{-n_\epsilon} < \epsilon.$$

Since $2^{-n-1} > 2^{-2n-2}$ when $n = 1, 2, \dots$, theorem 2.11 yields:

$$\phi\left(\frac{|g_{n+1}(x) - g_n(x)|}{2^{-n-1}}\right) = 0, \text{ almost everywhere (for each } n \geq n_\epsilon).$$

Next we observe that, for all x ,

$$\begin{aligned} \frac{|g(x) - g_n(x)|}{\epsilon} &\leq \frac{|g_{n+1}(x) - g_n(x)|}{2^{-n-1}} + \frac{|g_{n+2}(x) - g_{n+1}(x)|}{2^{-n-2}} + \dots \\ &\leq \sup_{p \geq n} \frac{|g_{p+1}(x) - g_p(x)|}{2^{-p-1}}, \text{ whenever } n \geq n_\epsilon. \end{aligned}$$

Since the union of a denumerable infinitude of sets of measure zero is likewise a set of measure zero, and since $\phi(u)$ is non-decreasing, we conclude that

$$\phi\left(\frac{|g(x) - g_n(x)|}{\epsilon}\right) = 0, \text{ almost everywhere (for each } n \geq n_\epsilon).$$

This implies that

$$l_{\phi, \psi}(g - g_n) \leq \epsilon, \text{ for all } n \geq n_\epsilon.$$

Case B. $\psi(\epsilon) > 0$, $\phi(0+) = 0$.

Here, determine n_ϵ so that

$$2T\phi\left(\frac{2^{-n_\epsilon}}{\epsilon}\right) \leq \psi(\epsilon).$$

Then, for all $n \geq n_\epsilon$,

$$\int \phi\left(\frac{|g(x) - g_n(x)|}{\epsilon}\right) dx \leq 2T\phi\left(\frac{2^{-n}}{\epsilon}\right) \leq \psi(\epsilon).$$

That is,

$$l_{\phi, \psi}(g - g_n) \leq \epsilon, \text{ for all } n \geq n_\epsilon.$$

Case C. $\psi(\epsilon) > 0$, $\phi(0+) > 0$.

Here, determine n_ϵ so that

$$2^{-n_\epsilon} < \min \left[\epsilon, \frac{\phi(0+)\psi(\epsilon)}{\psi(1)} \right].$$

By 2.13, with $u_0 = 0$,

$$\text{meas } \{|g_{n+1}(x) - g_n(x)| > 0\} \leq \frac{\psi(1)}{\phi(0+)} 2^{-n-1}, \quad (n = 1, 2, \dots).$$

Whence

$$\begin{aligned} \int \phi \left(\frac{|g(x) - g_n(x)|}{\epsilon} \right) dx &\leq \phi \left(\frac{2^{-n}}{\epsilon} \right) \text{meas} \{ |g(x) - g_n(x)| > 0 \} \\ &\leq \phi(1) \sum_{p=n}^{\infty} \text{meas} \{ |g_{p+1}(x) - g_p(x)| > 0 \} \\ &\leq \frac{\psi(1)}{\phi(0+)} 2^{-n} \leq \psi(\epsilon), \text{ for all } n \geq n_\epsilon. \end{aligned}$$

That is,

$$l_{\phi, \psi}(g - g_n) \leq \epsilon, \text{ for all } n \geq n_\epsilon.$$

Thus we have shown that the function $g(x)$ constructed at the outset of our proof meets the requirement:

$$\lim_{n \rightarrow \infty} l_{\phi, \psi}(g - g_n) = 0.$$

Since $g = (g - g_n) + g_n$ and since $L_{\phi, \psi}$ is linear we see that this same $g(x)$ is also an element of $L_{\phi, \psi}$.

2.4. On non-decreasing sequences of non-negative functions.

THEOREM 2.41. *If $h_1(x), h_2(x), \dots$ are elements of a given $L_{\phi, \psi}$, if for each j , ($j = 1, 2, \dots$).*

$$0 \leq h_j(x) \leq h_{j+1}(x), \text{ almost everywhere, } (-\infty < x < \infty),$$

and if there is a constant, M , independent of j , such that

$$l_{\phi, \psi}(h_j) \leq M, \quad (j = 1, 2, \dots),$$

then

$$h(x) = \lim_{j \rightarrow \infty} h_j(x) \leq \infty$$

exists almost everywhere, is an element of $L_{\phi, \psi}$, and is such that

$$l_{\phi, \psi}(h) \leq M.$$

Proof. Since the denumerably many inequalities, $h_j(x) \leq h_{j+1}(x)$, are valid simultaneously except at most on a set of measure zero, it follows that $h(x)$ exists almost everywhere.

Let $\epsilon > M$ be given. Determine some ϵ_0 such that $\epsilon > \epsilon_0 > M$. By 2.11,

$$\int \phi \left(\frac{h_j(x)}{\epsilon_0} \right) dx \leq \psi(\epsilon_0) \leq \psi(\epsilon), \quad (j = 1, 2, \dots).$$

As $j \rightarrow \infty$, the non-negative integrand increases monotonely. Hence,

$$\int \left\{ \lim_{j \rightarrow \infty} \phi \left(\frac{h_j(x)}{\epsilon_0} \right) \right\} dx = \lim_{j \rightarrow \infty} \int \phi \left(\frac{h_j(x)}{\epsilon_0} \right) dx.$$

But, by (1.11),

$$\lim_{j \rightarrow \infty} \phi \left(\frac{h_j(x)}{\epsilon_0} \right) \geq \phi \left(\frac{h(x)}{\epsilon} \right), \text{ almost everywhere.}$$

Whence,

$$\int \phi \left(\frac{h(x)}{\epsilon} \right) dx \leq \psi(\epsilon)$$

for each $\epsilon > M$. This implies that $l_{\phi, \psi}(h) \leq M$, and so establishes the theorem.

Note 2.42. Under the hypotheses of theorem 2.41 we might expect to conclude further that

$$\lim_{j \rightarrow \infty} l_{\phi, \psi}(h - h_j) = 0.$$

We shall see indeed, in theorem 3.22, that this conclusion can be guaranteed when $L_{\phi, \psi}$ has the property S defined in 3.11. However, if $L_{\phi, \psi}$ fails to have the property S , then the conclusion under consideration can be shown to be false by means of the functions constructed in theorem 3.41. The details will be discussed in note 3.23.

2.5. Proof that each $L_{\phi, \psi}$ is complete.

THEOREM 2.51. *If $f_1(x), f_2(x), \dots$ are elements of a given $L_{\phi, \psi}$ and if*

$$\lim_{m, n \rightarrow \infty} l_{\phi, \psi}(f_m - f_n) = 0,$$

then there exists an element $f(x)$ of $L_{\phi, \psi}$ such that

$$\lim_{n \rightarrow \infty} l_{\phi, \psi}(f - f_n) = 0.$$

Proof. If the theorem can be established for a sub-sequence of the given sequence of elements, then a simple application of 2.22 will establish the theorem for the given sequence. Therefore, we need prove our theorem in no more general a case than the following in which we assume that

$$l_{\phi, \psi}(f_m - f_n) \leq 2^{-2^{p-2}}, \text{ whenever } m \geq p \text{ and } n \geq p, \quad (p = 1, 2, \dots).$$

This implies, by 2.13 with $u_0 = 1$, that for each p , ($p = 1, 2, \dots$),

$$\text{meas} \{ |f_m(x) - f_n(x)| > 2^{-p-1} \} \leq \psi(1) 2^{-p-1}, \quad (m \geq p, n \geq p).$$

For each j , ($j = 1, 2, \dots$), let the corresponding set

$$\sum_{p=j}^{\infty} \{ |f_{p+1}(x) - f_p(x)| > 2^{-p-1} \}$$

be denoted by E_j and let the complement with respect to the entire interval, $-\infty < x < \infty$, be denoted by \bar{E}_j . We observe that

$$\text{meas } E_j \leq \sum_{p=j}^{\infty} \text{meas} \{ |f_{p+1}(x) - f_p(x)| > 2^{-p-1} \} \leq \psi(1) 2^{-j}, \quad (j = 1, 2, \dots),$$

and that each set E_j contains the corresponding set \bar{E}_{j+1} . Finally, we see that x is in \bar{E}_j if and only if *all* of the inequalities

$$|f_{p+1}(x) - f_p(x)| \leq 2^{-p-1}, \quad (p = j, j+1, \dots),$$

are valid.

Let us temporarily fix j . Construct the functions:

$$g_n(x) = \begin{cases} f_n(x), & \text{if, simultaneously, } x \text{ is in } \bar{E}_j \text{ and } |x| \leq j, \\ 0, & \text{otherwise,} \end{cases} \quad (n = 1, 2, \dots).$$

By 2.12, the sequence $g_1(x), g_2(x), \dots$ is seen to satisfy the hypotheses of 2.31 (with $T = j$). Hence,

$$g(x) = \lim_{n \rightarrow \infty} g_n(x)$$

exists for all x , is an element of $L_{\phi, \psi}$, and is such that

$$\lim_{n \rightarrow \infty} l_{\phi, \psi}(g - g_n) = 0.$$

Now, by 2.22 and 2.12,

$$\begin{aligned} l_{\phi, \psi}(g - g_n) &\leq l_{\phi, \psi}(g - g_m) + l_{\phi, \psi}(g_m - g_n) \\ &\leq l_{\phi, \psi}(g - g_n) + l_{\phi, \psi}(f_m - f_n) \\ &\leq l_{\phi, \psi}(g - g_m) + 2^{-2n-2}, \text{ for all } m \geq n. \end{aligned}$$

Thus,

$$l_{\phi, \psi}(g - g_n) \leq 2^{-2n-2}, \quad (n = 1, 2, \dots).$$

This last estimate is of importance in that $M = 2^{-2n-2}$ is independent of j . As $j \rightarrow \infty$, we see that $g(x)$ converges almost everywhere (since E_j contracts and since $\text{meas } E_j \rightarrow 0$) to a limiting function which we shall denote by $f(x)$. Moreover, since for each n , ($n = 1, 2, \dots$),

$$g(x) - g_n(x) = \begin{cases} f(x) - f_n(x), & \text{if, simultaneously, } x \text{ is in } \bar{E}_j \text{ and } |x| \leq j, \\ 0, & \text{otherwise,} \end{cases}$$

we find that if we fix n and allow $j \rightarrow \infty$, then the non-negative functions $|g(x) - g_n(x)|$ form a non-decreasing sequence converging almost everywhere to $|f(x) - f_n(x)|$. By 2.41, therefore,

$$l_{\phi, \psi}(f - f_n) \leq 2^{-2n-2}, \quad (n = 1, 2, \dots).$$

This proves that $f(x)$ is the desired element of $L_{\phi, \psi}$.

III. CONDITIONS FOR THE SEPARABILITY OF $L_{\phi, \psi}$.

3.1. The property S .

Definition 3.11. $L_{\phi, \psi}$ will be said to have the property S if the corresponding functions $\phi(u), \psi(u)$ satisfy the following additional requirements:

$$(3.111) \quad \phi(0+) = 0.$$

$$(3.112) \quad 0 < \psi(u) < \infty \text{ for all } u, 0 < u < \infty.$$

$$(3.113)^5 \quad \text{There exists a constant } C, 1 \leq C < \infty,$$

such that

$$\phi(2u) \leq C \phi(u), \quad (0 < u < \infty).$$

THEOREM 3.12. If $L_{\phi, \psi}$ satisfies (3.113), then

$$0 < \phi(u) < \infty \text{ whenever } 0 < u < \infty.$$

Proof. Since $\phi(1) = 1$, (3.113) yields:

$$0 < \phi(2^{-n}), \quad \phi(2^n) < \infty, \quad (n = 0, 1, 2, \dots).$$

Since $\phi(u)$ is non-decreasing, the theorem follows.

3.2. Dominated convergence in an $L_{\phi, \psi}$ having the property S .

THEOREM 3.21. If $f(x)$ is an element of a space $L_{\phi, \psi}$ having the property S , then $|f(x)| < \infty$, almost everywhere.

Proof. Apply (3.112) and 3.12.

THEOREM 3.22. If $g(x), f_1(x), f_2(x), \dots$ are elements of an $L_{\phi, \psi}$ with the property S , and if, almost everywhere,

$$\begin{cases} f(x) = \lim_{n \rightarrow \infty} f_n(x) \text{ exists,} \\ |f_n(x)| \leq |g(x)|, \end{cases} \quad (n = 1, 2, \dots),$$

then $f(x)$ is an element of $L_{\phi, \psi}$, and

$$\lim_{n \rightarrow \infty} l_{\phi, \psi}(f - f_n) = 0.$$

Proof. We may assume that

$$l = l_{\phi, \psi}(g) > 0,$$

as otherwise the theorem is trivial. Let $\epsilon > 0$ be given. Let us determine a corresponding non-negative integer $m = m_\epsilon$ such that

$$2^m > \frac{4l}{\epsilon}.$$

⁵ Compare this requirement that $\phi(u)$ be "not too convex" with the property (Δ_2) used by Birnbaum and Orlicz: "Über die Verallgemeinerung des Begriffes der zueinander konjugierten Potenzen," *Studia Mathematica*, t. 3 (1931), pp. 1-67.

Now $f(x)$, itself, is dominated almost everywhere by $g(x)$. Hence, by (3.113),

$$\phi\left(\frac{|f(x) - f_n(x)|}{\epsilon}\right) \leq \phi\left(\frac{2|g(x)|}{\epsilon}\right) \leq C^m \phi\left(\frac{2|g(x)|}{2^m \epsilon}\right) \leq C^m \phi\left(\frac{|g(x)|}{2l}\right),$$

almost everywhere. But, by 2.11 and (3.112),

$$\int C^m \phi\left(\frac{|g(x)|}{2l}\right) dx \leq C^m \psi(2l) < \infty.$$

Furthermore, by (3.111),

$$\lim_{n \rightarrow \infty} \phi\left(\frac{|f(x) - f_n(x)|}{\epsilon}\right) = 0, \text{ almost everywhere.}$$

By the Lebesgue theorem on dominated convergence, therefore,

$$\lim_{n \rightarrow \infty} \int \phi\left(\frac{|f(x) - f_n(x)|}{\epsilon}\right) dx = 0 < \psi(\epsilon).$$

In other words,

$$\limsup_{n \rightarrow \infty} l_{\phi, \psi}(f - f_n) \leq \epsilon.$$

This being true for each $\epsilon > 0$, our theorem is established.

Note 3.23. If we apply 3.22 by making the substitutions:

$$g(x) = h(x), \quad f_n(x) = h_j(x), \quad (n = j = 1, 2, \dots),$$

we learn, under the hypotheses of 2.41 and the additional requirement that $L_{\phi, \psi}$ have the property S , that

$$\lim_{j \rightarrow \infty} l_{\phi, \psi}(h - h_j) = 0.$$

Now let $L_{\phi, \psi}$ fail to have the property S . If we can show that the above extension of 2.41 is here false, then we shall have shown at the same time that 3.22 would be false without the hypothesis that $L_{\phi, \psi}$ have the property S . Consider the one-parameter family

$$\{f_\lambda(x)\}, \quad (0 \leq \lambda \leq 1),$$

which will be constructed in 3.41. Select, for example, the sequence,

$$\{1 - 1/j\}, \quad (j = 1, 2, \dots),$$

of parametric values and denote the corresponding elements of the family by $h_1(x)$, $h_2(x)$, \dots . These functions satisfy the hypotheses of 2.41 and converge for all x to the limiting function:

$$h(x) = \lim_{j \rightarrow \infty} h_j(x) = f_1(x).$$

But, by 3.41,

$$\liminf_{j \rightarrow \infty} l_{\phi, \psi}(h - h_j) \geq l > 0.$$

Thus, when 3.41 will have received its independent proof, the present necessity for the property S will have been established.

THEOREM 3.24. *If $f(x)$ is an element of an $L_{\phi, \psi}$ having the property S , and if for each T , ($T = 1, 2, \dots$), we set*

$$f^{(T)}(x) = \begin{cases} f(x), & \text{if, simultaneously, } |f(x)| \leq T \text{ and } |x| \leq T, \\ 0, & \text{otherwise,} \end{cases}$$

then each $f^{(T)}(x)$ is an element of $L_{\phi, \psi}$, and

$$\lim_{T \rightarrow \infty} l_{\phi, \psi}(f - f^{(T)}) = 0.$$

Proof. By 2.12, each $f^{(T)}(x)$ is in $L_{\phi, \psi}$. By 3.21,

$$\lim_{T \rightarrow \infty} f^{(T)}(x) = f(x), \text{ almost everywhere.}$$

By setting $g(x) = f(x)$, we obtain the present theorem from 3.22.

THEOREM 3.25. *If $f(x)$ is measurable, is bounded, and differs from zero at most on a finite interval, and if $L_{\phi, \psi}$ has the property S , then $f(x)$ is an element of $L_{\phi, \psi}$.*

Proof. Apply (3.112) and 3.12.

3.3. Proof that each $L_{\phi, \psi}$ with the property S is separable.

Definition 3.31. The expression "rational step-function" will be reserved for functions of the form:

$$s(x) = \begin{cases} a_k, & |x - k\delta| < \delta/2, \quad (k = 0, \pm 1, \dots, \pm n), \\ 0, & \text{otherwise,} \end{cases}$$

where n is a positive integer, where δ is a positive rational number, and where each a_k is a complex number whose real and imaginary parts are both rational. It is to be noted that the totality of rational step-functions can be arranged in a sequence:

$$s_1(x), \quad s_2(x), \quad \dots$$

THEOREM 3.32. *If $f(x)$ is an element of a given $L_{\phi, \psi}$ having the property S , and if $\epsilon > 0$, then there can be selected from the sequence of rational step-functions an element, $s_j(x)$, such that*

$$l_{\phi, \psi}(f - s_j) < \epsilon.$$

Proof. In virtue of 2.22 and 3.24, we need prove the theorem only in that case in which there exists a positive integer T such that

$$\begin{cases} |f(x)| \leq T, & \text{for all } x, \\ f(x) = 0, & \text{if } |x| > T. \end{cases}$$

Since $f(x)$ is also measurable, it is integrable. For each positive integer n let us determine the numbers:

$$b_{k,n} = n/T \int_{(k-\frac{1}{2})T/n}^{(k+\frac{1}{2})T/n} f(t) dt, \quad (k = 0, \pm 1, \dots, \pm n).$$

Let $\{a_{k,n}\}$ be any corresponding rational complex numbers such that

$$|a_{k,n}| \leq |b_{k,n}| \quad \text{and} \quad |a_{k,n} - b_{k,n}| < 1/n.$$

Consider the rational step-functions:

$$\sigma_n(x) = \begin{cases} a_{k,n}, & |x - kT/n| < T/2n, \quad (k = 0, \pm 1, \dots, \pm n), \\ 0, & \text{otherwise,} \end{cases} \quad (n = 1, 2, \dots).$$

Almost everywhere, the derivative of the indefinite integral of $f(x)$ exists and is equal to $f(x)$. Whence, readily,

$$\lim_{n \rightarrow \infty} \sigma_n(x) = f(x), \quad \text{almost everywhere.}$$

Next, we observe that

$$|\sigma_n(x)| \leq g(x) = \begin{cases} T, & |x| \leq T, \\ 0, & \text{otherwise,} \end{cases} \quad (n = 1, 2, \dots).$$

By 3.25, $L_{\phi,\psi}$ contains $g(x)$ and each $\sigma_n(x)$. Therefore, by 3.22,

$$\lim_{n \rightarrow \infty} l_{\phi,\psi}(f - \sigma_n) = 0.$$

This fact, since $\{\sigma_n(x)\}$ is a sub-sequence of $\{s_j(x)\}$, establishes the theorem.

3.4. Spaces $L_{\phi,\psi}$ lacking the property S .

THEOREM 3.41. *If $L_{\phi,\psi}$ fails to have the property S , then there exist numbers l and M , $0 < l < M < \infty$, and there exists a one-parameter family $\{f_\lambda(x)\}$, $(0 \leq \lambda \leq 1)$, of elements of $L_{\phi,\psi}$ such that*

$$l_{\phi,\psi}(f_\lambda) \leq M, \quad l_{\phi,\psi}(f_\lambda - f_\mu) \geq l, \quad (\lambda \neq \mu).$$

In particular, there exist such elements for which, furthermore,

$$0 \leq f_\lambda(x) \leq f_\mu(x), \quad \text{for all } x, \quad (\lambda < \mu).$$

Proof. Since, by 2.23, $l_{\phi, \psi}(f_\lambda - f_\mu) = l_{\phi, \psi}(f_\mu - f_\lambda)$, we shall introduce no loss of generality if we replace the requirement $\lambda \neq \mu$ by $\lambda < \mu$. Throughout the following four cases it will be agreed, firstly, that each definition or relation involving λ alone holds for all λ such that $0 \leq \lambda \leq 1$, and secondly, that each inequality involving both λ and μ holds whenever $0 \leq \lambda < \mu \leq 1$.

Case A. (3.111) is false.

Here, by 1.21, we can determine l , $0 < l < 1$, so that

$$0 \leq \psi(l) < \psi(1) < \infty.$$

And we can find a corresponding $M > 1$ such that

$$\phi\left(\frac{1}{M}\right)\psi(l) < \phi(0+)\psi(1).$$

Now, let

$$f_\lambda(x) = \begin{cases} \lambda, & 0 < x < \frac{\psi(1)}{\phi\left(\frac{1}{M}\right)}, \\ 0, & \text{otherwise.} \end{cases}$$

Since

$$\int \phi\left(\frac{f_\lambda(x)}{M}\right) dx \leq \psi(1) \leq \psi(M),$$

we see that

$$l_{\phi, \psi}(f_\lambda) \leq M.$$

And since

$$\int \phi\left(\frac{f_\mu(x) - f_\lambda(x)}{l}\right) dx \geq \phi(0+) \frac{\psi(1)}{\phi\left(\frac{1}{M}\right)} > \psi(l),$$

we conclude:

$$l_{\phi, \psi}(f_\mu - f_\lambda) \geq l.$$

Case B. (3.113) is false.

• Here, there exists a sequence: u_1, u_2, \dots , $0 < u_k < \infty$, such that

$$\phi(2u_k) > k\phi(u_k), \quad (k = 1, 2, \dots).$$

Determine corresponding positive numbers c_1, c_2, \dots so that

$$\sum_{k=1}^{\infty} c_k \phi(u_k) \leq \psi(1), \quad \sum_{k=1}^{\infty} c_k \phi(2u_k) = \infty.$$

We assert that the constants $l = 1$ and $M = 2$, and the functions:

$$f_\lambda(x) = \begin{cases} 2u_k, & 0 < x - \sum_{n=1}^{k-1} c_n < \lambda c_k, \\ 0, & \text{otherwise,} \end{cases} \quad (k = 1, 2, \dots),$$

meet the requirements of the theorem. Since

$$\int \phi\left(\frac{f_\lambda(x)}{2}\right) dx = \lambda \sum_{k=1}^{\infty} c_k \phi(u_k) \leq \psi(1) \leq \psi(2),$$

we see that

$$l_{\phi, \psi}(f_\lambda) \leq 2.$$

And since

$$\phi(f_\mu(x) - f_\lambda(x)) dx = \sum_{k=1}^{\infty} (\mu - \lambda) c_k \phi(2u_k) = \infty > \psi(1),$$

we conclude:

$$l_{\phi, \psi}(f_\mu - f_\lambda) \geq 1.$$

Case C. (3.113) is valid; (3.112) is false because there exists $l, 0 < l < 1$, such that $\psi(l) = 0$.

For this l and for $M = 1$, we assert that the functions:

$$f_\lambda(x) = \begin{cases} \lambda, & 0 < x < \psi(1), \\ 0, & \text{otherwise,} \end{cases}$$

meet the requirements of the theorem. Since $\phi(1) = 1$, we see that

$$\int \phi(f_\lambda(x)) dx \leq \psi(1),$$

and hence that

$$l_{\phi, \psi}(f_\lambda) \leq 1.$$

Since 3.12 yields:

$$\int \phi\left(\frac{f_\mu(x) - f_\lambda(x)}{l}\right) dx > 0 = \psi(l),$$

we conclude:

$$l_{\phi, \psi}(f_\mu - f_\lambda) \geq l.$$

Case D. (3.112) is false because there exists $M, 1 < M < \infty$, such that $\psi(M) = \infty$.

For $l = 1$ and for the above M , we assert that the functions:

$$f_\lambda(x) = \begin{cases} 1, & 0 < x - k < \lambda, \\ 0, & \text{otherwise,} \end{cases} \quad (k = 1, 2, \dots),$$

meet the requirements of the theorem. Since $\psi(M) = \infty$, we see that

$$l_{\phi, \psi}(f_\lambda) \leq M.$$

And since

$$\phi(f_\mu(x) - f_\lambda(x))dx = \sum_{k=1}^{\infty} (\mu - \lambda) = \infty > \psi(1),$$

we conclude:

$$l_{\phi, \psi}(f_\mu - f_\lambda) \geq 1.$$

THEOREM 3.42. *If $L_{\phi, \psi}$ is separable, then $L_{\phi, \psi}$ has the property S .*

Proof. Assume that $L_{\phi, \psi}$ lacks the property S , but that there exists a sequence $g_1(x), g_2(x), \dots$ dense in $L_{\phi, \psi}$. Consider the constant l and the family $\{f_\lambda(x)\}$ furnished by 3.41. For each $f_\lambda(x)$, determine a corresponding $g_j(x)$ such that

$$l_{\phi, \psi}(f_\lambda - g_j) < l/3,$$

and write: $f_\lambda \sim g_j$. Now, by 2.22,

$$l \leq l_{\phi, \psi}(f_\mu - f_\lambda) \leq 2l/3 + l_{\phi, \psi}(g_k - g_j),$$

whenever $f_\lambda \sim g_j$ and $f_\mu \sim g_k$. Thus, by 2.21, the correspondence between the non-denumerable family and the sequence is one-to-one. This contradiction establishes the theorem.

COLUMBIA UNIVERSITY.

THE ISOPERIMETRIC INEQUALITY ON THE SPHERE.¹

By TIBOR RADÓ.

1. *Statement and preliminary discussion of results. Reduction to a geometrical lemma.*

1.1. Among all simple closed plane curves with given length the circle has the greatest area. This classical theorem can also be expressed as follows: The area a of the circle with perimeter l is given by the equation

$$4\pi a = l^2.$$

Hence, on account of the extremal property of the circle, we have for every simple closed plane curve with length l and enclosed area a the *isoperimetric inequality in the plane*

$$4\pi a \leq l^2,$$

where the sign of equality holds if and only if the curve is a circle.

1.2. Similar considerations, applied to simple closed curves on the unit sphere, lead us to surmise the following *isoperimetric inequality on the unit sphere*: if l is the length and a the area enclosed by a simple closed curve on the unit sphere, then

$$(1) \quad 4\pi a - a^2 \leq l^2,$$

where the sign of equality holds if and only if the curve is a circle.

Since a simple closed curve determines *two* regions on the unit sphere, with areas a and $4\pi - a$ respectively, it is interesting to observe that the inequality (1) holds for *both* of these regions. Indeed, the function $\phi(x) = 4\pi x - x^2$ satisfies the equation

$$\phi(x) = \phi(4\pi - x),$$

and hence, if (1) holds for one of the two regions determined by a curve, then (1) also holds for the other region.

¹ Presented to the American Mathematical Society at the meeting in Chicago, April, 1935.

1.3. The inequality (1) has been proved by F. Bernstein for the important special case of curves which can be covered by a hemisphere,² and as far as the present author is aware, the general case has not been investigated as yet. It is the purpose of the present paper to establish (1) in complete generality.

1.4. In the sequel, C will denote a simple closed curve on the unit sphere, l the length of C , and a the area of any one of the two regions determined by C on the unit sphere. The following two cases exhaust then all possibilities.

Case I. The curve C can be covered by some open hemisphere.³

Case II. There exists no open hemisphere covering C .

1.5. Case I is settled by the result of F. Bernstein (see 1.3). While it seems that the ingenious method of F. Bernstein does not apply to case II, we shall see presently that this case is easily disposed of by means of the following lemma.

LEMMA. *If a simple closed curve C , located on the unit sphere, has a length less than 2π , then C can be covered by some open hemisphere.*

1.6. We shall prove the above lemma in § 2. If this result is anticipated, the discussion of case II in 1.4 is quite trivial. Indeed, if C cannot be covered by any open hemisphere, then the length l of C is $\geq 2\pi$, on account of the lemma. On the other hand, the maximum of the function

$$\phi(x) = 4\pi x - x^2$$

is

$$\phi(2\pi) = (2\pi)^2.$$

Hence (1) is certainly satisfied if $l \geq 2\pi$, and the sign of equality can hold if and only if $l = 2\pi$, $a = 2\pi$. But then C divides the surface of the unit sphere into two regions with equal areas, and consequently C contains two points P_1, P_2 which are end-points of the same diameter of the unit sphere.⁴ From $l = 2\pi$, in connection with the fact that the geodesics on the unit sphere are

² *Mathematische Annalen*, vol. 60 (1905), pp. 117-136.

³ An open hemisphere is a hemisphere *without* its boundary circle, and a closed hemisphere is a hemisphere *with* its boundary circle.

⁴ *Loc. cit.* ².

arcs of great circles, it follows then that C consists of two arcs of great circles joining P_1 and P_2 . On account of $a = 2\pi$ it follows finally that these arcs are sub-arcs of the same great circle; that is to say, that C itself is a great circle.

We proceed now to prove the lemma of 1. 5.

2. *Proof of the geometrical lemma.*

2. 1. In the sequel, C denotes a simple closed curve on the unit sphere, $l(C)$ denotes the length of C , and it is assumed that $l(C) < 2\pi$.

2. 2. Obviously, C cannot contain two points which are end-points of the same diameter of the unit sphere. Indeed, if P_1, P_2 were two such points, then both arcs determined on C by P_1, P_2 would have a length $\geq \pi$, and thus we should have $l(C) \geq 2\pi$, while by assumption $l(C) < 2\pi$.

2. 3. Let A_1, A_2, A_3 be three points of intersection of a great circle Γ with C . These points determine three non-overlapping arcs $\Gamma_1, \Gamma_2, \Gamma_3$ on Γ and three non-overlapping arcs C_1, C_2, C_3 on C , where Γ_1, C_1 are the arcs bounded by A_2, A_3 , and so on. Then one of the arcs $\Gamma_1, \Gamma_2, \Gamma_3$ has a length $> \pi$. Indeed, if we assume that

$$l(\Gamma_1) \leq \pi, l(\Gamma_2) \leq \pi, l(\Gamma_3) \leq \pi,$$

then $\Gamma_1, \Gamma_2, \Gamma_3$ are all three shortest arcs, on the unit sphere, between their respective end-points. Hence

$$l(C) = l(C_1) + l(C_2) + l(C_3) \geq l(\Gamma_1) + l(\Gamma_2) + l(\Gamma_3) = 2\pi,$$

in contradiction with the assumption that $l(C) < 2\pi$.

2. 4. Denote by S the set of all the points of intersection of a great circle Γ with C . In case S is empty, the conclusion to be stated at the end of the present section is obvious. Therefore we assume that S is not empty. Then S is a closed set, and the complementary set on Γ is the sum of non-overlapping open arcs $\gamma_1, \gamma_2, \dots$. Since the sum of the lengths of these complementary arcs is finite, namely $\leq 2\pi$, there is clearly one among them whose length is greater than or equal to the lengths of all the others. Suppose the notation is such that

$$l(\gamma_1) \geq l(\gamma_n), \quad (n = 1, 2, 3, \dots).$$

Denote by P_1, Q_1 the end-points of γ_1 . Then P_1, Q_1 are points of S and consequently of C , and hence (see 2.2) P_1 and Q_1 are not end-points of the same diameter of Γ . Consequently, $l(\gamma_1) \neq \pi$. We wish to show that $l(\gamma_1) > \pi$. Suppose, indeed, that

$$(2) \quad l(\gamma_1) < \pi.$$

Denote by P_1^*, Q_1^* the points of Γ diametrically opposite to P_1, Q_1 , respectively, and by γ_1^* the sub-arc of Γ which is bounded by P_1^*, Q_1^* , and which does not overlap with γ_1 . On account of the assumption (2), such a sub-arc γ_1^* does exist (see figure) and we have

$$l(\gamma_1^*) = l(\gamma_1).$$

On account of 2.2, the points P_1^*, Q_1^* are not in the set S . Since S is closed, we have therefore vicinities of P_1^* and Q_1^* which are free of points of S . Hence, if we first assume that γ_1^* does not contain any point of S , then there follows the existence of an open sub-arc $\bar{\gamma}$ of Γ which is free of points of S and which is longer than γ_1^* . Such an arc $\bar{\gamma}$ would then be comprised in one of the complementary arcs $\gamma_1, \gamma_2, \dots$, say in γ_k , and we should have

$$l(\gamma_k) \geq l(\bar{\gamma}) > l(\gamma_1^*) = l(\gamma_1).$$

This contradicts however the extremal property of γ_1 . Secondly, let us assume that γ_1^* contains some point R^* of S . Then R^* is different both from P_1^* and Q_1^* , since these points are not in S . Hence (see figure) none of the three non-overlapping arcs, determined by P_1, Q_1, R^* on Γ , has a length $> \pi$. This contradicts however 2.3. Summing up:

On every great circle Γ there exists an open sub-arc with a length $> \pi$ which is free of points of C .

2.5. It will be convenient to use the following definitions. A point-set Σ , located on the unit sphere, will be said to possess the property *CH* if it can be covered by some closed hemisphere. The set Σ will be said to possess the property *OH* if it can be covered by some open hemisphere.

2.6. Let Σ be a closed sub-arc of the curve C (the case $\Sigma \equiv C$ is not excluded). If Σ possesses the property *CH*, then it also possesses the property *OH*.

For the sake of brevity of presentation, the closed hemisphere which covers Σ will be called the southern hemisphere, and the boundary of this hemisphere will be called the equator E . On account of 2.4, we have then on the equator E an open sub-arc $\bar{\gamma}$, free of points of C and consequently of points of Σ , such that $l(\bar{\gamma}) > \pi$. We can choose then on $\bar{\gamma}$ a closed sub-arc γ such that $l(\gamma) = \pi$. Denote by A, B the end-points of γ . Then A and B are the end-points of the same diameter of the unit sphere. Since γ and Σ are closed sets without common points, a slight rotation around the diameter AB will carry γ into an arc γ^* which has no points in common with Σ . Let us denote by E^* the transform of the equator E under such a rotation. Let us choose the sense of rotation in such a way that γ^* is located on the closed southern hemisphere. Then the open arc $E^* - \gamma^*$ is located on the *open* northern hemisphere, and has consequently no points in common with Σ which is located on the *closed* southern hemisphere. Since γ^* also has no common points with Σ , it follows that the great circle $E^* = \gamma^* + (E^* - \gamma^*)$ does not intersect the arc Σ . Hence Σ is located on one of the two open hemispheres bounded by E^* , that is to say Σ has the property OH .

2.7. Let us now choose a sense on the curve C which we shall call the positive sense, and let us also choose a point A on C . We shall denote by $P(s)$ the point of C at distance s from A , the distance being measured on C in the positive sense from A to $P(s)$. Then s varies in the closed interval

$$(3) \quad 0 \leq s \leq l(C) < 2\pi.$$

2.8. Let us denote by $\{s; CH\}$ the set of all those values of s , in the interval (3), for which the following assertion is true: the closed sub-arc $\gamma(s)$ of C , consisting of all points $P(\sigma)$ such that $0 \leq \sigma \leq s$, possesses the property CH . The set $\{s; CH\}$ is not empty, since sufficiently small values of s are certainly comprised in $\{s; CH\}$. Furthermore, $\{s; CH\}$ is obviously a closed set, and contains therefore a largest number \bar{s} . We assert that $\bar{s} = l(C)$. Suppose, indeed, that $\bar{s} < l(C)$. Then the arc $\gamma(\bar{s})$ is a true sub-arc of C . Since \bar{s} is in $\{s; CH\}$, the arc $\gamma(\bar{s})$ possesses the property CH , and therefore, on account of 2.6, the property OH also. That is to say, the closed arc $\gamma(\bar{s})$ can be covered by some open hemisphere. By obvious reasons of continuity, the same open hemisphere will also cover the arc $\gamma(\bar{s} + \epsilon)$, provided $\epsilon > 0$ is sufficiently small. Hence, for $\epsilon > 0$ sufficiently small, the arc $\gamma(\bar{s} + \epsilon)$ possesses the property OH and consequently the property CH . This contradicts however the extremal property of \bar{s} . Hence $\bar{s} = l(C)$, that is to say the whole curve C possesses the property CH and therefore, on account of 2.6 the property OH also. Thus the lemma of 1.5 is proved.

Conclusion. The following general problem arise with the topics considered in the present paper. Let surface S , subject only to the usual restrictions of differential geometry. Denote by a the area and by l the of a simply connected region on S . What is the best such regions on S , for a in terms of l ? The reader in references indicating the scope of the results obtained

If the surface S is a plane, then the classical $4\pi a \leq l^2$ solves the problem. If S is the unit sphere $4\pi a - a^2 \leq l^2$, as discussed in the present paper. If then Carleman proved that the answer is $4\pi a \leq l^2$, then (T. Carleman, "Zur Theorie der Minimalflächen," *M* vol. 9 (1921), pp. 154-160). If S is a saddle-shape if the Gaussian curvature of S is ≤ 0 throughout, then $4\pi a \leq l^2$. And conversely, if the inequality $4\pi a \leq l^2$ a surface, S , then the Gaussian curvature of S is necessarily results are due to E. F. Beckenbach and T. Radó, 'and surfaces of negative curvature," *Transactions of the Mathematical Society*, vol. 35, pp. 662-674). Obviously problems of this nature are yet to be solved.

THE OHIO STATE UNIVERSITY,
COLUMBUS, OHIO.

THE CONJUGATE CHORD QUADRICS OF A CURVE ON A SURFACE.

By M. L. MACQUEEN.

1. *Introduction.* The projective differential geometry of a surface in ordinary space has been greatly enriched by the knowledge of a number of quadrics that have been associated with a point of a surface. Among these quadrics there are the two *asymptotic osculating quadrics* of Bompiani;¹ each of these is the limit of the quadric determined by three asymptotic tangents of one family constructed at points of a curve C on a surface S at these points independently approach P along C .

Lane² has defined in a similar way two quadrics called *conjugate osculating quadrics*, replacing, however, in the definition just stated, the three asymptotic tangents by three consecutive tangents of the curves of one family of a given conjugate net.

Another pair of quadrics called *asymptotic chord quadrics* has been defined³ by Bompiani, these quadrics having a geometrical property in common with the asymptotic osculating quadrics, namely, that they contain the asymptotic tangents at the point P on the surface S . An asymptotic chord quadric may be defined as follows: On a surface S let us select a curve C and two points P and \bar{P} on this curve. Denote the points of intersection of the two asymptotic curves through each of these points by P_1 and P_2 . As \bar{P} approaches P along C the asymptotic chord $\bar{P}P_1$ generates a ruled surface which has for every position of its generators a well defined osculating quadric. The limiting position of this quadric as $\bar{P} \rightarrow P$ along C is called an asymptotic chord quadric. A second quadric may be defined similarly by considering the asymptotic chord $\bar{P}P_2$.

It would seem to be of interest to define similarly two quadrics called *conjugate chord quadrics*, which are associated with each point of a given

¹ E. Bompiani, "Ancora sulla geometria delle superficie considerate nello spazio rigato," *Rendiconti dei Lincei* (6), vol. 4 (1926), p. 262.

² E. P. Lane, "Conjugate nets and the lines of curvature," *American Journal of Mathematics*, vol. 53 (1931), pp. 573-588. Hereinafter referred to as Lane, "Conjugate Nets."

³ E. Bompiani, "Sugli elementi di 2° ordine delle curve di una superficie," *Rendiconti dei Lincei* (6), vol. 9 (1929), p. 288.

curve on a surface sustaining a given conjugate net. For this purpose we shall replace the asymptotic curves through P and \bar{P} , in the definition of an asymptotic chord quadric, by the curves of each family of a given conjugate net.

In § 2 we summarize briefly for subsequent use some of the theory of the projective differential geometry of a surface in ordinary space referred to a conjugate net as parametric. In § 3, after formulating a definition of the conjugate chord quadrics, the equations of these quadrics are found. The method which we employ is quite similar to that used by Bompiani in obtaining the equations of the asymptotic chord quadrics. Finally, in § 4, we study briefly the geometrical properties of these quadrics, our investigation paralleling somewhat that of the conjugate osculating quadrics made by Lane.

2. *Analytic basis.* In this section an analytic basis for the projective differential geometry of a parametric conjugate net on a surface in ordinary space is established. We employ the completely integrable system of differential equations, which are written in a symmetrical form intimately associated with the axis congruence of the net, that define a conjugate net in ordinary space, except for a projective transformation.

Let the projective homogeneous coördinates $x^{(1)}, \dots, x^{(4)}$ of a point P_x on a surface S referred to a conjugate net N_x in ordinary space be given as analytic functions of two independent variables u, v . The two osculating planes of the parametric curves C_u, C_v at a point P_x on the surface S intersect in a line defined to be the axis of P_x with respect to the net N_x . If P_y is any point distinct from P_x on the axis and if P_y is chosen such that it is the harmonic conjugate of P_x with respect to the two foci of the axis regarded as generating a congruence when the point P_x varies over the surface S , then x and y satisfy a system of differential equations⁴ of the form

$$(1) \quad \begin{aligned} x_{uu} &= px + \alpha x_u + Ly, \\ x_{uv} &= cx + ax_u + bx_v, \\ x_{vv} &= qx + \delta x_v + Ny \quad (LN \neq 0), \end{aligned}$$

where the notation here employed is similar to that used by Lane in his recent book.

The coefficients of system (1) are functions of u, v and satisfy four integrability conditions which we do not need to write here.

From the equations

⁴ E. P. Lane, *Projective Differential Geometry of Curves and Surfaces*, The University of Chicago Press, 1932, p. 138.

$$(x_{vv})_u = (x_{uv})_v, \quad (x_{uu})_v = (x_{uv})_u$$

we find

$$(2) \quad \begin{aligned} y_u &= fx - nx_u + sx_v + Ay, \\ y_v &= gx + tx_u + nx_v + By, \end{aligned}$$

where we have placed

$$(3) \quad \begin{aligned} fN &= c_v + ac + bq - c\delta - q_u, & gL &= c_u + bc + ap - c\alpha - p_v, \\ -nN &= a_v + a^2 - a\delta - q, & tL &= a_u + ab + c - \alpha_v, \\ sN &= b_v + ab + c - \delta_u, & nL &= b_u + b^2 - b\alpha - p, \\ A &= b - (\log N)_u, & B &= a - (\log L)_v. \end{aligned}$$

The Laplace-Darboux point invariants H, K , the tangential invariants \mathbf{H}, \mathbf{K} , the invariants $\mathfrak{V}, \mathfrak{C}', \mathfrak{D}$ of Green, and the invariant r are given in the notation of system (1) by the formulas

$$(4) \quad \begin{aligned} H &= c + ab - a_u, & K &= c + ab - b_v, \\ \mathbf{H} &= sN, & \mathbf{K} &= tL, \\ 8\mathfrak{V} &= 4a - 2\delta + (\log r)_v, & 8\mathfrak{C}' &= 4b - 2\alpha - (\log r)_u, \\ \mathfrak{D} &= -2nL, & r &= N/L. \end{aligned}$$

The differential equation of the asymptotic curves on the surface S is

$$(5) \quad L du^2 + N dv^2 = 0.$$

Since the four points x, x_u, x_v, y are not coplanar, we shall choose them as vertices of a local tetrahedron of reference at P_x . Let the unit point be chosen so that a point X defined by

$$(6) \quad X = x_1x + x_2x_u + x_3x_v + x_4y$$

has local coördinates proportional to x_1, \dots, x_4 . Later we shall have occasion to use the covariant tetrahedron with vertices at the points x, ρ, σ, y for which

$$(7) \quad \sigma = x_v - ax, \quad \rho = x_u - bx.$$

The points σ, ρ are the ray-points, or Laplace transformed points, of the curves C_u, C_v respectively, corresponding to the point P_x . •

3. *The conjugate chord quadrics.* In order to formulate a definition let us consider a surface S referred to a conjugate net N_x , and any curve C on S not belonging to N_x . Let P_x and \bar{P} be two distinct points of the curve C . Let $P_1(P_2)$ be the point of intersection of the u -curve (v -curve) through P_x and the v -curve (u -curve) through \bar{P} . As the point \bar{P} approaches P_x along C ,

the line joining \bar{P} to P_1 , namely, the conjugate chord surface which has for every position of one of its generators an osculating quadric. The limit of this quadric as $\bar{P} \rightarrow P_x$ is a *conjugate chord quadric* at the point P_x of the curve C . A second conjugate chord quadric is defined in a way similar to the first by considering the conjugate chord $\bar{P}P_2$.

In order to find the equations of the conjugate chord surface let us consider a curve C through points P_x and \bar{P} on C with parametric equations

$$(8) \quad u = u(t), \quad v = v(t).$$

If the points P_x and \bar{P} of the curve C correspond to the parameters $t = t_1$ and $t = t_2$, and if the coördinates of points \bar{P} , P_1 , and P_2 are \bar{x} , x_1 , and x_2 respectively, we have

$$(9) \quad \begin{aligned} \bar{x} &= x + A_1 t + A_2 t^2/2 + A_3 t^3/6 + \dots \\ x_1 &= x + R_1 t + R_2 t^2/2 + R_3 t^3/6 + \dots \\ x_2 &= x + S_1 t + S_2 t^2/2 + S_3 t^3/6 + \dots \end{aligned}$$

where

$$(10) \quad \begin{aligned} A_1 &= x_u u' + x_v v', & A_2 &= x_{uu} u'^2 + 2x_{uv} u'v' + x_{vv} v'^2 \\ A_3 &= x_{uuu} u'^3 + 3x_{uuv} u'^2 v' + 3x_{uvv} u'v'^2 + x_{vvv} v'^3 \\ &\quad + 3x_{uv}(u'v'' + u''v') + 3x_{vv}v'v'' + x_u u''' \end{aligned}$$

and

$$(11) \quad \begin{aligned} R_1 &= x_u u', & R_2 &= x_{uu} u'^2 + x_u u'', & R_3 &= x_{uuu} u'^3 \\ S_1 &= x_v v', & S_2 &= x_{vv} v'^2 + x_v v'', & S_3 &= x_{vvv} v'^3 \end{aligned}$$

The formulas for the third derivatives of x expressed in terms of x , x_u , x_v , y are found from system (1) by differentiation to be

$$(12) \quad \begin{aligned} x_{uuu} &= (p_u + \alpha p + fL)x + (\alpha_u + \alpha^2 + p - nL)x_u \\ &\quad + sLx_v + (L_u + \alpha L + AL)x_v \\ x_{uuv} &= (c_u + bc + ap)x + (a_u + c + ab + \alpha\alpha)x_u \\ &\quad + (b_u + b^2)x_v + aLy, \\ x_{uvv} &= (c_v + ac + bq)x + (a_v + a^2)x_u \\ &\quad + (b_v + c + ab + b\delta)x_v - \delta y \\ x_{vvv} &= (q_v + \delta q + gN)x + tNx_u \\ &\quad + (\delta_v + \delta^2 + g + nN)x_v - \delta y \end{aligned}$$

It is evident that the points X , Y defined by

$$(13) \quad X = (\bar{x} + x_1)/2, \quad Y = (\bar{x} - x_1)/2$$

are on the conjugate chord $\bar{P}P_1$. Any point P_ϕ , except the point Y , on the line $\bar{P}P_1$ is defined by placing

$$(14) \quad \phi(t, \mu) = X(t) + \mu Y(t) \quad (\mu \text{ scalar}).$$

If t is fixed while μ varies, the locus of the point P_ϕ is the line $\bar{P}P_1$. If t, μ both vary, the locus of P_ϕ is the ruled surface of the conjugate chords of one family (v -curves) relative to the curve C .

Fixing a generator of the ruled surface by holding t fixed, we observe that an asymptotic tangent at a point of the fixed generator of the ruled surface, distinct from a rectilinear generator, at the point (t, μ) joins the point P_ϕ to the point $d\phi$, if $dt/d\mu$ is determined by the condition

$$(15) \quad |\phi, \phi_\mu, \phi_t, \phi_{tt}dt + 2\phi_{t\mu}d\mu| = 0.$$

If we vary μ , with $dt/d\mu$ determined by (15), t being fixed, the line joining ϕ and $d\phi$ describes an osculating quadric along a fixed generator. When t approaches zero we obtain the desired quadric at P_x .

It is easy to verify that condition (15) may be written in the form

$$(16) \quad |X, Y, X' + \mu Y', \phi_{tt}| dt + 2 |X, Y, X', Y'| d\mu = 0,$$

wherein the accents denote derivatives with respect to t . When $t = 0$, easy calculations yield the following results:

$$(17) \quad \begin{aligned} X &= x, & Y &= x_v v', & X' &= x_u u' + x_v v'/2, \\ Y' &= x_{uv} u' v' + x_{vv} v'^2/2 + x_v v''/2, \\ X'' &= x_{uu} u'^2 + x_{uv} u' v' + x_{vv} v'^2/2 + x_u u'' + x_v v''/2, \\ Y'' &= x_{uv} u'^2 v' + x_{uvv} u' v'^2 + x_{vvv} v'^3/3 + x_{uv}(u' v'' + u'' v') \\ &\quad + x_{vv} v' v'' + x_v v'''/3. \end{aligned}$$

On substituting in equation (16) the expression for ϕ_{tt} found from equation (14), and reducing by means of elementary properties of determinants, we find that the first determinant appearing in (16) may be written in the form

$$(18) \quad |X, Y, X' + \mu Y', \phi_{tt}| = |X, Y, X' + \mu Y', X'' + \mu Y''| \\ = D + E\mu + F\mu^2,$$

where

$$(19) \quad \begin{aligned} D &= |X, Y, X', X''|, & F &= |X, Y, Y', Y''|, \\ E &= |X, Y, X', Y''| + |X, Y, Y', X''|. \end{aligned}$$

On making use of (17), (1), and (12) we find

$$\begin{aligned}
|X, Y, X', Y'| &= -N\Delta u'v'^3/2, \\
D &= -(Lu'^2 + Nv'^2/2)\Delta u'v', \\
E &= \{N(\alpha u'^2 + au'v' + u'')v'/2 - a(Lu'^2 + Nv'^2/2)u' \\
(20) \quad &- [aLu'^2 + bNu'v' + (N_v + \delta N + BN)v'^2/3 + Nv'']u'\}\Delta v'^2, \\
F &= \{N[(a_u + c + ab + a\alpha)u'^2v' + (a_v + a^2)u'v'^2 + tNv'^3/3 \\
&+ a(u'v'' + u''v')]/2 \\
&- au'[aLu'^2 + bNu'v' + (N_v + \delta N + BN)v'^2/3 + Nv'']\}\Delta v'^3,
\end{aligned}$$

where

$$\Delta = |x, x_u, x_v, y| \neq 0.$$

By means of the results just obtained it is now possible to write equation (16) in the form

$$(21) \quad (\bar{D} + \bar{E}\mu + \bar{F}\mu^2)dt = Nu'v'^2d\mu,$$

where we have placed

$$\bar{D} = D/\Delta v', \quad \bar{E} = E/\Delta v', \quad \bar{F} = F/\Delta v'.$$

The quadric whose equation we are seeking is generated by the point

$$(22) \quad \zeta = k\phi + d\phi = k\phi + \phi_t dt + \phi_\mu d\mu \quad (k \text{ scalar}),$$

as μ and k vary and if $dt/d\mu$ is determined by (21). On making use of (14), (17), and (21), we find that ζ can be expressed in the form

$$\begin{aligned}
(23) \quad \zeta &= k(x + \mu x_v v') + Nu'v'^2[(x_u u' + x_v v')/2] \\
&+ \mu(x_{uv}u'v' + x_{vv}v'^2/2 + x_v v''/2)] + (\bar{D} + \bar{E}\mu + \bar{F}\mu^2)x_v v'.
\end{aligned}$$

Then by use of (1) we find

$$\zeta = x_1 x + x_2 x_u + x_3 x_v + x_4 y,$$

where

$$\begin{aligned}
x_1 &= k + N\mu u'v'^3(cu' + qv'/2), \\
x_2 &= Nu'^2v'^2(1 + a_\mu v'), \\
(24) \quad x_3 &= k\mu v' + Nu'v'^3/2 + N\mu u'v'^3(2bu' + \delta v' + v''/v')/2 \\
&+ (\bar{D} + \bar{E}\mu + \bar{F}\mu^2)v', \\
x_4 &= N^2\mu u'v'^4/2.
\end{aligned}$$

These are the local coördinates x_1, \dots, x_4 of the point P_ζ , referred to the tetrahedron x, x_u, x_v, y with suitably chosen unit point. Homogeneous elimination of μ and k from equations (24), together with some further simplification, the details of which will be omitted, gives the equation of the conjugate chord quadric Q_v , referred to the tetrahedron x, x_u, x_v, y , namely,

$$(25) \quad L\lambda x_2^2 + N\lambda^2 x_2 x_3 - 2a\lambda x_3 x_4 - 2\lambda x_1 x_4 + Px_4^2 + Qx_2 x_4 = 0,$$

wherein we have placed $\lambda = dv/du$, and

$$P = 2[(c + ab - a_u) + (q + a\delta - a_v - a^2)\lambda - iN\lambda^2/3]/N, \\ Q = 2(N_v/N - \delta/2 + B)\lambda^2/3 - a\lambda + \lambda(\log \lambda)'.$$

In order to write the equation of the quadric Q_v referred to the tetrahedron x, ρ, σ, y , it is not difficult to show that it is sufficient to replace x_1 in equation (25) by $x_1 - bx_2 - ax_3$. On making this substitution and simplifying the coefficients by means of (3) and (4) we find the *equation of the conjugate chord quadric Q_v referred to the covariant tetrahedron x, ρ, σ, y to be*

$$(26) \quad Lx_2^2 + [4(\mathfrak{C}' + \lambda\mathfrak{B}'/3) + (\log \lambda r^{1/2})']x_2 x_4 \\ - [\mathfrak{D}/L - 2H/\lambda N + 2\lambda\mathbf{K}/3L]x_4^2 - 2x_1 x_4 + \lambda Nx_2 x_3 = 0.$$

The equation of the conjugate chord quadric Q_u at the point P_x of the curve C on the net N_x can be written immediately by interchanging u and v and making the necessary symmetrical interchanges of the other symbols. For this result we find

$$(27) \quad N\lambda x_3^2 + [4(\mathfrak{C}'/3 + \lambda\mathfrak{B}') - \lambda(\log \lambda r^{1/2})']x_3 x_4 \\ + [\lambda\mathfrak{D}/L + 2\lambda^2\mathbf{K}/L - 2\mathbf{H}/3N]x_4^2 - 2\lambda x_1 x_4 + Lx_2 x_3 = 0.$$

4. *Geometrical considerations.* Let us regard the curve C , employed in defining the conjugate chord quadrics, as imbedded in the one-parameter family of curves defined on the surface S by the equation

$$(28) \quad dv - \lambda du = 0,$$

where λ is a non-vanishing function of u, v . Then the differential equation

$$(29) \quad (dv - \lambda du)(dv - \mu du) = 0$$

will represent a conjugate net on the surface S if the two directions defined by this equation separate harmonically the two asymptotic directions satisfying equation (5). A necessary and sufficient condition for this is

$$(30) \quad \mu = -1/\lambda r.$$

The two curves of such a conjugate net that pass through a point on the surface S will be denoted by C_λ and C_μ respectively, according as the direction dv/du has the value λ or μ .

The tangent line of the curve C_λ at the point P_x is given by

$$(31) \quad x_4 = x_3 - \lambda x_2 = 0,$$

and that of the curve C_μ is

$$(32) \quad x_4 = x_2 + \lambda r x_3 = 0.$$

Inspection of equation (27) shows that the conjugate chord quadric Q_u is intersected by the tangent plane, $x_4 = 0$, of the surface S in the u -tangent, $x_4 = x_3 = 0$, and in a residual line which is precisely the tangent line of the curve C_μ given by (32). Similarly, the tangent plane intersects the quadric Q_v in the v -tangent, $x_4 = x_2 = 0$, and in the same residual line (32). Therefore, *the conjugate chord quadric Q_u (Q_v) at every point of a curve C on a conjugate net intersects the tangent plane of the net in the u -tangent (v -tangent) and in a residual line tangent to the curve in the direction conjugate to the curve C .*

It is well known that a necessary and sufficient condition for a direction to coincide with its conjugate is that it be an asymptotic direction. Hence, *the residual line in which the tangent plane of the net intersects the two conjugate chord quadrics at every point of a curve on a net coincides with the tangent to the curve if, and only if, the curve is an asymptotic curve on the sustaining surface.*

Lane has shown ⁵ that the conjugate osculating quadric Q_u is intersected by the tangent plane in the u -tangent and in the residual line whose equation, referred to the same tetrahedron that we are using, is

$$(33) \quad x_4 = (L - N\lambda^2)x_3 - 2\lambda Lx_2 = 0.$$

Similarly, the tangent plane intersects the conjugate osculating quadric Q_v in the v -tangent and in a residual line in general distinct from the line (33). The two residual lines in which the tangent plane intersects the two conjugate osculating quadrics coincide ⁶ if, and only if, the curve C_λ is an asymptotic curve on S , and in this case the lines coincide with the tangent of the curve.

The residual line (32) of the conjugate chord quadric Q_u coincides with the residual line (33) of the conjugate osculating quadric Q_u if, and only if, $1 - r\lambda^2 = 0$, i. e., if C_λ is an asymptotic curve on S . Therefore, *the conjugate chord quadrics and the conjugate osculating quadrics at every point of a curve on a net are intersected by the tangent plane of the net in the same lines if, and only if, the curve is an asymptotic curve on the surface; then the lines are the u -tangent, the v -tangent, and the tangent of the curve.*

⁵ Lane, "Conjugaté nets," p. 579.

⁶ *Ibid.*, p. 580.

We next consider briefly the intersections of the conjugate chord quadrics with the two osculating planes $x_3 = 0$, $x_2 = 0$, of the curves C_u , C_v respectively. One of the generators of the quadric Q_u , in the osculating plane $x_3 = 0$ at the point P_x of the curve C_u , is the u -tangent, $x_3 = x_4 = 0$. Another generator in this plane is the line

$$(34) \quad (\mathfrak{D} + 2\lambda K - 2\mathbf{H}/3\lambda r)x_4 - 2Lx_1 = 0.$$

In the osculating plane $x_3 = 0$, one of the generators of the conjugate osculating quadric Q_u is found ⁷ to be the u -tangent also, and the line

$$(35) \quad (\mathfrak{D} + \lambda K - \mathbf{H}/\lambda r)x_4 - 2Lx_1 = 0.$$

Thus, on comparing equations (34) and (35) we find that *the osculating plane $x_3 = 0$, at a point P_x of a curve C_u , contains the tangent of this curve as a common generator of the conjugate chord quadric Q_u and the conjugate osculating quadric Q_u at a point P_x of a curve C_λ . The other generators, in the plane $x_3 = 0$, of these two quadrics coincide if, and only if,*

$$\lambda^2 = -\mathbf{H}/3rK.$$

As is also true in the case of the conjugate osculating quadric Q_u , the osculating plane, $x_3 = 0$, touches the conjugate chord quadric Q_u of the curve C_λ in the ray-point ρ of the curve C_v .

Results similar to the preceding may be obtained by considering the quadric Q_v and the osculating plane, $x_2 = 0$, of the curve C_v .

The equation of the cone projecting from the point P_x the curve of intersection of the two conjugate chord quadrics Q_u , Q_v is found by eliminating x_1 from their equations. For this result we find

$$(36) \quad \begin{aligned} & L(x_3 - \lambda x_2)(x_2 + r\lambda x_3) + [4(\mathfrak{G}'/3 + \lambda \mathfrak{B}') - \lambda(\log \lambda r^{1/2})']x_3x_4 \\ & - \lambda[4(\mathfrak{G}' + \lambda \mathfrak{B}'/3) + (\log \lambda r^{1/2})']x_2x_4 \\ & + 2[\lambda \mathfrak{D}/L - (H + \mathbf{H}/3)/N + \lambda^2(K + \mathbf{K}/3)/L]x_4^2 = 0. \end{aligned}$$

This cone will be tangent to the tangent plane, $x_4 = 0$, along the tangent line (31) of the curve C_λ in case $1 + r\lambda^2 = 0$, that is, in case the curve C_λ is an asymptotic curve.

It is of some interest to consider the relations of the two conjugate chord quadrics Q_u for the curves C_λ , C_μ of the two conjugate families. The equation of the quadric Q_u for the curve C_μ , with μ determined by equation (30), can be written immediately by replacing λ by $-1/r\lambda$ in equation (27). This result is found to be

⁷ *Ibid.*, p. 580.

$$(37) \quad Nx_3^2 + [4(\mathfrak{B}' - r\lambda\mathfrak{C}'/3) + (\log \lambda r^{1/2})_u - (\log \lambda r^{1/2})_v/r\lambda]x_3x_4 \\ - 2[n + K/\lambda N - \lambda H/3L]x_4^2 - \lambda Nx_2x_3 - 2x_1x_4 = 0.$$

If x_1 is eliminated from equations (27) and (37), we obtain the equation of the cone projecting from the point P_x the curve of intersection of the conjugate chord quadrics Q_u for the curves of the two conjugate families, namely,

$$(38) \quad [4(1 + r\lambda^2)r\mathfrak{C}'/3 - 2r\lambda(\log \lambda r^{1/2})_u + (1 - r\lambda^2)(\log \lambda r^{1/2})_v]x_3x_4 \\ + N(1 + r\lambda^2)x_2x_3 + 2(1 + r\lambda^2)(K - H/3)x_4^2/L = 0.$$

This cone is indeterminate in case $1 + r\lambda^2 = 0$, that is, when the curve C_λ is an asymptotic curve. In this case the two quadrics (27) and (37) coincide. In case $1 + r\lambda^2 \neq 0$ and $H = 3K$ the cone (38) is a pair of planes so that the two quadrics intersect in two conics. One of the planes is the osculating plane, $x_3 = 0$, of the curve C_u , and the equation of the other is easily read from equation (38). Otherwise the cone (38) is tangent to the osculating plane, $x_3 = 0$, of the curve C_u , touching it along its tangent line, $x_3 = x_4 = 0$.

SOUTHWESTERN COLLEGE,
MEMPHIS, TENN.

ARITHMETICAL THEOREMS ON LUCAS FUNCTIONS AND TCHEBYCHEFF POLYNOMIALS.

By E. T. BELL.

1. *Introduction and Summary.* Any identity in elliptic, or elliptic theta functions, is equivalent to (implies and is implied by) an identity in Lucas¹ functions or Tchebycheff polynomials with arbitrary complex arguments. We first develop an isomorphism between the Lucas functions and circular functions.

2. *The functions U, V, C, S, T .* Principal values of all irrationalities, wherever they occur, are to be understood. Let α, β, z, u be complex variables such that $\alpha\beta(\alpha - \beta) \neq 0$. Write $b \equiv \alpha + \beta$, $c \equiv \alpha\beta$. Then α, β are the roots of

$$(1) \quad \begin{aligned} y^2 - by + c &= 0; \\ 2\alpha &= b + (b^2 - 4c)^{1/2}, & 2\beta &= b - (b^2 - 4c)^{1/2}. \end{aligned}$$

The functions U, V , associated with (1), are defined by

$$(2) \quad U(b, c, z) \equiv \frac{\alpha^z - \beta^z}{\alpha - \beta}, \quad V(b, c, z) \equiv \alpha^z + \beta^z.$$

The connection with circular functions mentioned in § 1 will be made by means of

$$(3) \quad C(u, z) \equiv \frac{1}{2}V(2u, 1, z), \quad S(u, z) \equiv U(2u, 1, z).$$

Thus $C(u, z), S(u, z)$ are associated with

$$(4) \quad y^2 - 2uy + 1 = 0,$$

of which the roots are

$$\rho = u + (u^2 - 1)^{1/2}, \quad \sigma = u - (u^2 - 1)^{1/2},$$

and we have

$$C(u, z) = \frac{1}{2}(\rho^z + \sigma^z), \quad S(u, z) = \frac{\rho^z - \sigma^z}{\rho - \sigma}.$$

Comparing the last with (1), (2), we have the important formulas

¹ The functions are the U_n, V_n of Lucas, *American Journal of Mathematics*, vol. 1 (1878), pp. 184-240, 288-321; the arguments are his P, Q , which he restricts to be rational. The designation "Tchebycheff Polynomials" is retained as it appears to be accepted, although these polynomials occur in the writings of Vieta, Newton, De Moivre, Johann Bernoulli and Euler, to mention only predecessors of Gauss and Cauchy. According to Lucas, *loc. cit.*, p. 208, they first appear in Vieta's *Opera*, Leyden, 1646, pp. 295-299.

$$(5) \quad U(b, c, z) = c^{(z-1)/2} S(b/2c^{1/2}, z), \quad V(b, c, z) = 2c^{z/2} C(b/2c^{1/2}, z).$$

To connect our functions with one usual definition (for example that of Polya and Szegö) of the Tchebycheff polynomials we introduce

$$(6) \quad T(u, z) \equiv S(u, z+1),$$

which, for z an integer, is a Tchebycheff polynomial. Only one of S , T is necessary, as any formula involving either may be restated immediately in terms of the other by means of (6). We shall use S in preference to T .

From (2) we have

$$(7) \quad U(b, c, 0) = 0, \quad U(b, c, 1) = 1; \quad V(b, c, 0) = 2, \quad V(b, c, 1) = b;$$

$$(8) \quad U(b, c, -z) = -c^{-z} U(b, c, z); \quad V(b, c, -z) = c^{-z} V(b, c, z);$$

$$(9) \quad W(b, c, z+n+2) - bW(b, c, z+n+1) + cW(b, c, z+n) = 0, \\ W = U, V; \quad n = 0, \pm 1, \pm 2, \dots$$

From (7), (9) the $U(b, c, z+n)$, $V(b, c, z+n)$ can be calculated by recurrence. From these and (3), (4) we write down the corresponding relations for C , S :

$$(10) \quad C(u, 0) = 1, \quad C(u, 1) = u; \quad S(u, 0) = 0, \quad S(u, 1) = 1;$$

$$(11) \quad C(u, -z) = C(u, z); \quad S(u, -z) = -S(u, z);$$

$$(12) \quad Z(u, z+n+2) - 2uZ(u, z+n+1) + Z(u, z+n) = 0, \\ Z = C, S; \quad n = 0, \pm 1, \pm 2, \dots$$

Let m be an integer > 0 . Then, identically,

$$\cos(m+2)\theta - 2\cos\theta\cos(m+1)\theta + \cos m\theta = 0, \\ \sin(m+2)\theta - 2\cos\theta\sin(m+1)\theta + \sin m\theta = 0.$$

Comparing the first of these with (10), (12), we see that $C(u, m)$ is the Tchebycheff polynomial obtained from the expansion of $\cos m\theta$ in powers of $\cos\theta$ on replacing $\cos\theta$ by u . From the second, $S(u, m+1)$ is the Tchebycheff polynomial $T(u, m)$ obtained on replacing $\cos\theta$ by u in the expansion of $\sin(m+1)\theta \csc\theta$ in powers of $\cos\theta$. Hence, for integer values of $m > 0$, we have the explicit forms

$$(13) \quad C(u, m) = m \sum_{r=0}^{[(m-1)/2]} \frac{(-1)^r}{m-2r} \binom{m-r-1}{r} 2^{m-2r-1} u^{m-2r},$$

$$(14) \quad S(u, m) = u \sum_{r=0}^{[(m-1)/2]} (-1)^r \binom{m-r-1}{r} 2^{m-2r-1} u^{m-2r},$$

where $[x]$ is the greatest integer in x . These are in descending powers of u . In ascending powers we have

$$(15) \quad C(u, 2m) = (-1)^m m \sum_{s=0}^m \frac{(-1)^s 2^{2s}}{m+s} \binom{m+s}{2s} u^{2s},$$

$$(16) \quad C(u, 2m-1) = (-1)^{m-1} (2m-1) u \sum_{s=0}^{m-1} \frac{(-1)^s 2^{2s}}{m+s} \binom{m+s}{2s+1} u^{2s},$$

$$(17) \quad S(u, 2m) = 2(-1)^{m+1} u \sum_{s=0}^{m-1} (-1)^s 2^{2s} \binom{m+s}{2s+1} u^{2s},$$

$$(18) \quad S(u, 2m-1) = (-1)^{m-1} \sum_{s=0}^{m-1} (-1)^s 2^{2s} \binom{m+s-1}{2s} u^{2s}.$$

To write down the expressions of $S(u, n)$, $C(u, n)$, where n is an arbitrary integer (positive, zero, or negative) we refer to (10), (11) and use the sgn function, defined for all real values of x by $\text{sgn } x = -1, 0, 1$ according as $x < 0$, $x = 0$, $x > 0$. Then

$$(19) \quad C(u, n) = C(u, |n|), \quad S(u, n) = \text{sgn } n S(u, |n|).$$

From (13)-(18) and (5) we have the following reduced forms of the $U(b, c, m)$, $V(b, c, m)$ for all integers $m > 0$:

$$(20) \quad V(b, c, m) = \sum_{r=0}^{[(m-1)/2]} \frac{(-1)^r m}{m-2r} \binom{m-r-1}{r} b^{m-2r} c^r,$$

$$(21) \quad U(b, c, m) = b \sum_{r=0}^{[(m-1)/2]} (-1)^r \binom{m-r-1}{r} b^{m-2r} c^r;$$

$$(22) \quad V(b, c, 2m) = 2(-1)^m m \sum_{s=0}^m \frac{(-1)^s}{m+s} \binom{m+s}{2s} b^{2s} c^{m-s},$$

$$(23) \quad V(b, c, 2m-1) = (-1)^{m-1} (2m-1) b \sum_{s=0}^{m-1} \frac{(-1)^s}{m+s} \binom{m+s}{2s+1} b^{2s} c^{m-s-1},$$

$$(24) \quad U(b, c, 2m) = (-1)^{m+1} b \sum_{s=0}^{m-1} (-1)^s \binom{m+s}{2s+1} b^{2s} c^{m-s-1},$$

$$(25) \quad U(b, c, 2m-1) = (-1)^{m-1} \sum_{s=0}^{m-1} (-1)^s \binom{m+s-1}{2s} b^{2s} c^{m-s-1}.$$

Corresponding to (19), for arbitrary integers n ,

$$(26) \quad \begin{aligned} V(b, c, n) &= c^{n(\text{sgn } n-1)/2} V(b, c, |n|), \\ U(b, c, n) &= \text{sgn } n c^{n(\text{sgn } n-1)/2} U(b, c, |n|), \end{aligned}$$

which refer all cases to (20)-(26). As (22)-(25) are merely (20), (21) written in reverse order, the numerical coefficients are integers.

We shall call $U(b, c, n)$, $V(b, c, n)$ the Lucas functions of the arguments b, c , of order n . Lucas had occasion to consider his functions only for rational (for the most part integral) arguments. The generating functions of $U(b, c, n)$,

$V(b, c, n)$ can be written down from (2), if needed; the information given by them is, however, more readily obtainable from the isomorphism next noted.

3. *Trigonometric identities.* Lucas (*loc. cit.*, p. 189) emphasized the connection between his functions and the circular (or hyperbolic) functions, and he gave a method for passing from identities in his functions to identities in circular functions. The method is also reversible. We shall give another way of accomplishing these things, which seems more direct and which is much easier to apply. What follows is generalized in § 5.

Let the n_i, r_j be integers, and let $P(z_1, \dots, z_{s+t})$ be a polynomial in z_1, \dots, z_{s+t} . For $0 < x < \pi/2$ let

$$P(\cos n_1 x, \dots, \cos n_s x, \sin r_1 x, \dots, \sin r_t x) = 0$$

be an identity in x . Replace $\sin r_j x$ by $\sin x \sin r_j x \csc x$ ($j = 1, \dots, t$). The result is an identity in x for the same range. In this we write

$$\begin{aligned} \cos n_i x &= C(u, n_i), & \sin r_j x \csc x &= S(u, r_j), \\ u = \cos x, & \sin x = (1 - u^2)^{1/2} & (i = 1, \dots, s; j = 1, \dots, t); \end{aligned}$$

say the result is $R(u) = 0$. Then $R(u) = 0$ is an identity in u , since $R(u)$ is a polynomial in $u, (1 - u^2)^{1/2}$ which vanishes for an infinity of distinct values of u . Hence in $R(u) = 0$ we may replace u by the complex variable w , getting the identity $R(w) = 0$ in w . Choose $w = b/2c^{1/2}$ (as in (5)). It follows that we can pass directly from the original \cos, \sin identity $P(\) = 0$ to $P(\dots, 2c^{-n_i/2}V(b, c, n_i), \dots, \frac{1}{2}(4c - b^2)^{1/2}c^{-r_j/2}U(b, c, r_j), \dots) = 0$. This converts the trigonometric identity $P(\) = 0$ into an identity in functions U, V with the arbitrary arguments b, c . The process is evidently reversible, and we may sum up the simple algorithm in the following table, in which n is an integer. (See also § 5.)

<i>Circular</i>	<i>Lucas</i>
$\cos nx$	$\frac{1}{2}c^{-n/2}V(b, c, n)$
$\sin nx \csc x$	$c^{(1-n)/2}U(b, c, n)$
$\sin x$	$(1 - b^2/4c)^{1/2}$
$\cos x$	$b/2c^{1/2}$

To pass from a trigonometric identity in x to its correspondent in Lucas functions we replace the functions in the first column by their correspondents in the second. To pass from an identity in Lucas functions to its correspondent in circular functions, we replace $V(b, c, n)$ by $2c^{n/2} \cos nx$, and $U(b, c, n)$ by $c^{(n-1)/2} \sin nx \csc x$. The last two lines of the table suffice to eliminate b, c

from the result, if they are not already absent; if not, the supposed identity is untrue.

We give some simple examples. All admit of generalization. From (m, n are integers)

$$\sin (m+n)x = \sin mx \cos nx + \cos mx \sin nx,$$

and the table, we have

$$2U(b, c, m+n) = U(b, c, m)V(b, c, n) + V(b, c, m)U(b, c, n).$$

Similarly

$$\cos (m+n)x = \cos mx \cos nx - \sin^2 x \frac{\sin mx}{\sin x} \frac{\sin nx}{\sin x},$$

$$2V(b, c, m+n) = V(b, c, m)V(b, c, n) + (b^2 - 4c)U(b, c, m)U(b, c, n).$$

To indicate the inverse process we verify the evident identity

$$U(b, c, 2n) = U(b, c, n)V(b, c, n),$$

which is obtained from the first of the above by taking $m = n$. If the identity is true, then

$$c^{(2n-1)/2} \sin 2nx \csc x = c^{(2n-1)/2} \sin nx \csc x 2c^{n/2} \cos nx,$$

which is $\sin 2nx = 2 \sin nx \cos nx$, and hence the identity is verified, since the steps are evidently reversible.

The algorithm provides a straightforward way for deriving relations between the U, V . Suppose, for example, we wish to sum

$$\sum_{s=1}^n V^p(b, c, s)$$

in terms of functions U, V , where p is a positive integer. By the table, this is equivalent to summing

$$2^p \sum_{s=1}^n (c^{p/2})^s \cos^p sx,$$

expressing the result as a function of sines and cosines, and reapplying the table to replace \sin, \cos by U, V . The special cases of this considered by Lucas follow in this way at once.

The isomorphism expressed in the table also suggests a possible generalization of the U, V which it might be profitable to investigate: the two functions U, V are replaced by the four polynomials in $\operatorname{sn} x, \operatorname{cn} x$ occurring in the numerators and denominators of $\operatorname{sn} nx, \operatorname{cn} nx, \operatorname{dn} nx$. When the modulus is zero, the polynomials degenerate to the U, V .

The isomorphism also leads to the theory of "division" for the U , V , which Lucas mentions in passing, but on which he did not, apparently, publish anything. This concerns the expression of $U(b, c, t)$, $V(b, c, t)$ with t rational but not integral, as algebraic functions (in the usual technical sense as in complex variables) of U 's and V 's of integral order. As this does not concern the application to elliptic functions which we have in view here, we shall not go into it.

If the identities involve several sets of variables x_i , b_i , c_i and associated integers n_{ij} , the table is applied to each. There is an example in § 4.

4. *Elliptic and theta identities.* It was shown in a previous paper² that any identity in elliptic or elliptic theta functions is equivalent to an identity in arbitrary parity functions summed over one or more quadratic partitions of integers. The parity functions may be replaced by products of circular functions (sines or cosines) having the same parity. This identity is also equivalent to the original identity. Transforming the circular identity by means of the table in § 3 into its equivalent in U , V functions, we get an identity in Lucas functions which is equivalent to the elliptic identity. We need give only enough examples³ to illustrate the process.

From the expansion of $\vartheta_1'/\vartheta_0^2(x)$ in arithmetical form, as given in M § 10, we have

$$\vartheta_1'^2/\vartheta_0^2(x) = 4\Sigma q^{m/2}[\Sigma t \cos(t - \tau)x] \\ (m = 1, 3, 5, \dots; m = t\tau, t > 0, \tau > 0).$$

Hence, using

$$\vartheta_0(x) = \Sigma q^{\nu^2}(-1)^\nu \cos 2\nu x \quad (\nu = 0, \pm 1, \pm 2, \dots),$$

multiplying throughout by $\vartheta_0^2(x)$, and applying

$$\vartheta_1'^2 = [\Sigma(-1)^{(\mu-1)/2} \mu q^{\mu^2/4}]^2, \quad (\mu = \pm 1, \pm 3, \pm 5, \dots),$$

we have the following identity on equating coefficients of like powers of q :

$$\Sigma t \cos(t - \tau + 2\nu_1 + 2\nu_2)x = \Sigma(-1)^{(m_1 m_2 - 1)/2} m_1 m_2,$$

where the sums refer to all positive integers t , τ , m_1 , m_2 , and to all integers ν_1 , ν_2 (positive, zero, or negative) such that, for n a constant integer ≥ 0 ,

$$4n + 2 = m_1^2 + m_2^2 = t\tau + 2\nu_1^2 + 2\nu_2^2.$$

² *Transactions of the American Mathematical Society*, vol. 22 (1921), p. 1; cited as A.

³ The expansions used are quoted either from the paper cited in the preceding footnote, or from my paper in *Messenger of Mathematics*, vol. 54 (1924), pp. 166-176, cited as M.

Applying the table in § 3, we have

$$\Sigma t c^{(\tau-t-2\nu_1-2\nu_2)} V(b, c, t - \tau + 2\nu_1 + 2\nu_2) = 2\Sigma (-1)^{(m_1 m_2 - 1)/2} m_1 m_2,$$

where the sums are as just defined and b, c are independent complex variables. The other identity of this sort is obtained from $\vartheta_1'^2/\vartheta_2'^2(x)$, by applying the method of A to the expansion in M § 10.

As a somewhat simpler specimen of the infinity of results obtainable from the available expansions we shall state the equivalent of the identity $[\vartheta_1'/\vartheta_3(x)]\vartheta_3(x) = \vartheta_1'$, using the expansion of $\vartheta_1'/\vartheta_1(x)$ in M § 8.

Let $(-1 | m)$, $(2 | m)$ be defined as Jacobi-Legendre symbols when m is an odd integer, and write $\epsilon(n) = 1$, or 0, according as n is, or is not, the square of an integer > 0 . Consider all the integer solutions, for α constant and $\equiv 1 \pmod{4}$, of

$$\alpha = t\tau + 4\nu^2 \quad (t > 0, \tau > 0, \nu \geq 0),$$

and let Σ refer to all sets of solutions (t, τ, ν) . Then

$$4\Sigma (-1 | \tau) (2 | t\tau) c^{(\tau-t-4\nu)/2} V(b, c, t - \tau + 4\nu) = \epsilon(m) (-1 | m^{1/2}) m^{1/2}.$$

Such identities can be given a form analogous to that of the recurrences for functions of divisors, of which the first example is due to Euler. Thus, from M § 8, we have

$$\vartheta_1'/\vartheta_0(x) = 2\Sigma q^{\alpha/4} \left[\Sigma (-1 | \tau) \cos \left(\frac{t-\tau}{2} x \right) \right],$$

where the first Σ refers to $\alpha = 1, 5, 9, 13, \dots$, and the second to all positive integers t, τ such that $\alpha = t\tau$. Write

$$P(b, c, \alpha) \equiv \Sigma (-1 | \tau) c^{(\tau-t)/2} V \left(b, c, \frac{t-\tau}{2} \right),$$

where Σ refers to t, τ , and $t\tau = \alpha$ (as before). Then the identity $[\vartheta_1'/\vartheta_0(x)]\vartheta_0(x) = \vartheta_1'$ is equivalent to

$$2\Sigma (-c)^{-\nu} P(b, c, m - 4\nu^2) = \epsilon(m) (-1 | m^{1/2}) m^{1/2},$$

where m is positive and $\equiv 1 \pmod{4}$, and Σ refers to $\nu = 0, \pm 1, \pm 2, \dots$, the sum continuing so long as $4\nu^2 < m$.

To illustrate identities in more than one variable we take $f(x, y) \equiv \cos \xi x \cos \eta y$ in a theorem in A,

$$2\Sigma [f(d_1 - d_2, \delta_1 + \delta_2) - f(d_1 + d_2, \delta_1 - \delta_2)] = \Sigma t [f(0, 2t) - f(2t, 0)]$$

in which $f(x, y)$ is an arbitrary single-valued function of x, y satisfying the conditions $f(x, y) = f(-x, y) = f(x, -y)$, and the sums refer to all positive

integers $d_1, \delta_1, d_2, \delta_2, t, \tau$ such that $2t\tau = d_1\delta_1 + d_2\delta_2$, $\tau, d_1, \delta_1, d_2, \delta_2$ are odd, and $t\tau$ is constant. The equivalent in Lucas functions is

$$\begin{aligned} & \Sigma[y_1^{(d_2-d_1)/2}y_2^{-(\delta_1+\delta_2)/2}V(x_1, y_1, d_2-\delta_2)V(x_2, y_2, \delta_1+\delta_2) \\ & \quad - y_1^{-(d_1+d_2)/2}y_2^{-(\delta_1-\delta_2)/2}V(x_1, y_1, d_2+d_2)V(x_2, y_2, \delta_1-\delta_2)] \\ & = 2\Sigma[y_2^{-t}V(x_2, y_2, 2t) - y_1^{-t}V(x_1, y_1, 2t)], \end{aligned}$$

5. *Generalized isomorphism.* Let x be a real or complex variable, and z a complex variable. Then the isomorphism may be generalized as follows, in which \sim is the sign of 1, 1 correspondence:

$$\begin{aligned} \cos x & \sim \tfrac{1}{2}c^{-z/2}V(z), & \sin x & \sim \tfrac{1}{2}c^{-z/2}\Delta^{1/2}U(z), \\ V(z) & \equiv V(b, c, z), & U(z) & \equiv U(b, c, z), & \Delta & \equiv \Delta(b, c) \equiv 4c - b^2, \end{aligned}$$

in which b, c as before are complex variables. For, the variables α, β in a U, V relation may be replaced by other variables, or by functions of themselves. Under the particular transformation $\alpha \rightarrow \alpha^z, \beta \rightarrow \beta^z$, we have, from the original isomorphism,

$$\cos nx \sim \tfrac{1}{2}c^{-nz/2}V(nz), \quad \sin nx \sim \tfrac{1}{2}c^{-nz/2}\Delta^{1/2}(nz),$$

as may be readily verified, in which n is an arbitrary integer. Taking $n = 1$ we get the form stated.

Let x_1, \dots, x_s be independent real or complex variables, and z_1, \dots, z_s independent complex variables. Then, more generally, we have the above isomorphism in which now

$$x \equiv x_1 + \dots + x_s, \quad z \equiv z_1 + \dots + z_s,$$

and the correspondence between the two sets of independent variables is $x_j \sim z_j$ ($j = 1, \dots, s$). This is evident upon mathematical induction on s applied to the addition theorems for \cos, \sin, V, U , and an application of the isomorphism as stated at the beginning of this section. Finally, in the same way, or as a consequence, we have the isomorphism in which

$$x \equiv n_1x_1 + \dots + n_sx_s, \quad z \equiv n_1z_1 + \dots + n_sz_s,$$

where the n 's are integers, and $n_jx_j \sim n_jz_j$ ($j = 1, \dots, s$).

BINARY QUADRATIC DISCRIMINANTS DIFFERING BY SQUARE FACTORS.

By GORDON PALL.

1. The problem of finding the number of representations of numbers not prime to the discriminant in an integral quadratic form, or in a genus or order thereof, has never been adequately solved. The classical methods depending on forms constructed from the roots of certain congruences, subject to conditions which isolate the forms of a given class, genus, or order, are simple enough for numbers prime to the discriminant, but are excessively detailed and tedious in general. One reason for these complications is that those methods lead first to proper representations, and through them to all, and the relations between proper representations in different classes and genera are more involved than those for all representations. To avoid the difficulty we may attempt to reduce the problem to one in which the order-invariants are (so far as possible) free of square factors.

We treat in this paper binary quadratic forms, our principal method employing reduction formulae to replace a problem concerning forms of discriminant p^2d by a like problem for discriminant d .

A discriminant d is a non-zero integer congruent to 0 or 1 (mod 4), and is fundamental if divisible by no prime p such that d/p^2 is a discriminant, i. e.

$$(1) \quad p > 2 \text{ and } p^2 \mid d, \text{ or } p = 2 \text{ and } d \equiv 0 \text{ or } 4 \pmod{16}.$$

Let $r(n, d)$ denote the number of sets¹ of representations of n in a representative system of primitive integral binary quadratic forms of discriminant d . Then $r(n, d)$ is a factorable function of n , and if p does not satisfy (1),

$$(2) \quad r(p^a, d) = 1 + (d \mid p) + (d \mid p^2) + \cdots + (d \mid p^a).^2$$

It is also known that

$$(3) \quad r(p, d) = 0 \text{ if } p \text{ satisfies (1).}^2$$

Consequently a prime is represented in at most one class (and the reciprocal class) of discriminant d . On the basis of this property and the composition of classes the writer has elsewhere (Pall I and II) considered the

¹ Pall II, 491; or Dirichlet, 216.

² Pall II, p. 493.

number $f(n)$ of sets of representations of n in any integral binary quadratic form f , but only partially for n divisible by primes of type (1).

To any primitive f_1 of discriminant p^2d corresponds a form f of discriminant d such that

$$(4) \quad f_1(p^2n) = \sigma f(n), \text{ for every integer } n.^3$$

Here $\sigma = \sigma_{p,d}$ is defined as follows:

$$(5) \quad \sigma = 1 \text{ if } d < -4 \text{ or } d \text{ is square, } \sigma = 2 \text{ if } d = -4, \sigma = 3 \text{ if } d = -3;$$

and if d is positive but not square, σ is the least positive integer such that $p \mid u_\sigma$, where (t_k, u_k) denote the successive integral solutions of

$$(6) \quad t^2 - du^2 = 4.$$

In § 5 we shall associate (4) with the Lipschitz correspondence between the classes of discriminants d and p^2d .⁴ The formula

$$(7) \quad h(p^2d)/h(d) = \{p - (d \mid p)\}/\sigma,$$

which was the object of Lipschitz's investigation, can be deduced directly from (4) and (9), $h(d)$ denoting the number of primitive classes of discriminant d .⁵

The slight variant of Lipschitz's method, given in § 2, may be of value. Lemma 3 replaces the reference to Gauss' complicated Art. 162, and finds further applications in § 3.

A primitive ambiguous class of discriminant d has ν derived primitive ambiguous classes of discriminant p^2d , where ν is given in Theorem 4. We note in § 3 the following curious corollary:

Frequently, an ambiguous class C of even discriminant d contains one ambiguous form $[a, b, c]$ with b/a even, and another with b/a odd. If $\sigma = \sigma_{2,d} > 1$ and $d \equiv 0, 12$, or $28 \pmod{32}$ such forms of both kinds seem to exist in every primitive ambiguous class; if $d \equiv 4$ or $8 \pmod{16}$, or $16 \pmod{32}$, then $\sigma = 1$, and it is trivial that every primitive ambiguous form has b/a even. But:

- (8) if $d \equiv 0, 12$, or $28 \pmod{32}$ and $\sigma = 1$, every primitive ambiguous class contains ambiguous forms of only one kind, and there are equally many ambiguous classes of either kind.

³ Pall I, p. 331, II, p. 494.

⁴ Lipschitz, p. 255.

⁵ Pall III.

Further properties of derived classes appear in section 4.

Finally we note in section 6 that

$$(9) \quad r(p^2n, p^2d) = \{p - (d | p)\}r(n, d),$$

which together with (2) and (3) yields immediately the value of $r(p^2, d)$ for any p .

If n has no factor of type (1), n is represented in at most one primitive genus of discriminant d . In any case it is represented equally often in each of the 2^t genera which represent it, specified in § 7.

2. Binary matrices P and Q are called *right-equivalent* if $PU = Q$ for some unitary integral matrix U ; *left-equivalent* if $UP = Q$.

$$(10) \quad \text{If } P = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, P^\dagger \text{ denotes the adjoint-transpose } \begin{pmatrix} \delta & -\beta \\ -\gamma & \alpha \end{pmatrix}.$$

Evidently $PU = Q$ is equivalent to $U^\dagger P^\dagger = Q^\dagger$, and every P is the adjoint-transpose of the matrix P^\dagger of the same determinant. We have

LEMMA 1. Every integral matrix of prime determinant p is right-equivalent to one and only one of the $p + 1$ matrices

$$(11) \quad P_\kappa = \begin{pmatrix} p & \kappa \\ 0 & 1 \end{pmatrix} \quad (\kappa = 0, 1, \dots, p-1), \quad P_p = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix};^6$$

and left-equivalent to one and only one of

$$(12) \quad P_\kappa^\dagger = \begin{pmatrix} 1 & -\kappa \\ 0 & p \end{pmatrix} \quad (\kappa = 0, 1, \dots, p-1), \quad P_p^\dagger = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}.$$

By fP we denote the form derived from $f = [a, b, c]$ by applying the linear transformation of matrix P . By the associativity of matrix multiplication, $f(PU) = fPU$. Hence

THEOREM 1. Let $f = [a, b, c]$ be primitive, $b^2 - 4ac = d \neq 0$. Every form derivable from f by integral transformations of prime determinant p is equivalent to one of the $p + 1$ forms

$$(13) \quad \begin{aligned} fP_\kappa &= [ap^2, (2a\kappa + b)p, a\kappa^2 + b\kappa + c], & (\kappa = 0, \dots, p-1), \\ fP_p &= [a, bp, cp^2]. \end{aligned}$$

Exactly $p - (d | p)$ of the forms (13) are primitive.

⁶ Mathews, p. 160; or Smith, pp. 166, 176.

The $1 + (d | p)$ imprimitive forms fP_κ are given

$$(14) \quad a\kappa^2 + b\kappa + c \equiv 0 \pmod{p} \quad (0 \leq \kappa < p)$$

if $p \nmid a$; by $\kappa = p$ and (14) if $p | a$. In every case fP_κ is primitive unless (1) holds, and then p^2 .

Let V be an unitary integral matrix. If $VP_\lambda = VP_\kappa U$, $P_\kappa = P_\lambda U$, $\kappa = \lambda$. Hence V is equivalent in some order to P_0, \dots, P_p . This proves

LEMMA 2. If f is replaced by an equivalent form gP_0, \dots, gP_p constitute a permutation of the classes.

LEMMA 3. Let f and g be integral primitive forms of determinant p and P and Q integral binary matrices of determinant p and primitive. Then $R = (PQ^\dagger)/p$ is integral, determinant 1.

For choose integral unitary matrices U and V such that $U^\dagger P = P_\lambda^\dagger$ and $Q^\dagger V = P_\kappa$, and set $f' = fU$, $g' = gV$, $R' = U^\dagger R V$, $f'R' = g'$, f' and g' are integral and $f'P = f'U^\dagger P = f'P_\lambda^\dagger$, $g'Q = g'V^\dagger Q = g'P_\kappa^\dagger$. It remains to show that $f'P_\lambda^\dagger$ is imprimitive. Now according to (1) $\kappa = p > \lambda$, (2) $\kappa < p = \lambda$, (3) $\kappa = \lambda = p$, R' is

$$(15) \quad \begin{pmatrix} 1 & (\kappa - \lambda)/p \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1/p & -\lambda \\ 0 & p \end{pmatrix}, \begin{pmatrix} p & \kappa \\ 0 & 1/p \end{pmatrix}$$

Let $f' = [a, b, c]$. In cases (1) and (2),

$$f'P_\lambda^\dagger = [a, -2a\lambda + bp, a\lambda^2 - b p \lambda]$$

and is imprimitive if $p | a$. If R' is not integral in case (3) $f'R' = [a, \dots, a\mu^2/p^2 + b\mu/p + c]$ is integral only if $p^2 | a$. The same holds for $f'E' = [a/p^2, \dots, \dots]$. In case (2) $f'R'$ is imprimitive if $p | c$, $f'R' = [ap^2, \dots, a\kappa^2 + b\kappa/p + c]$ if $p | c$.

THEOREM 2. Let d be a discriminant, p a prime not dividing d .

The conclusion may be false if $fP = gQ$ is imprimitive unless (1) holds, and all we can say is that f and g are derived from the same discriminant d/p^2 . If the divisor is p , p is represented in the discriminant d ; then the classes of f and g are either identical or distinct; this may be deduced from

$$fP_p/p = [a/p, b, cp] \sim [cp, -b, a/p] = [c, -b, a]$$

class K of discriminant p^2d is derivable from a primitive class C of discriminant d by an integral transformation of determinant p . (ii) C is uniquely determined by K and p .

(i) Replace d by p^2d in Theorem 1. The imprimitive form fP_κ is unique, and of divisor p^2 ; set $fP_\kappa = p^2f'$, whence f' is imprimitive and of discriminant d . Then

$$(16) \quad p^2f = f(P_\kappa P_\kappa^\dagger) = (fP_\kappa)P_\kappa^\dagger = p^2f'P_\kappa^\dagger, \quad f = f'P_\kappa^\dagger.$$

(ii) The uniqueness follows from Lemma 3.

The proof of the fact that each of the $p - (d | p)$ primitive forms (13) is equivalent to exactly $\sigma - 1$ of the others, is based on the following criterion. If $fP \sim fQ$, $fPU = fQ$ for some unitary integral matrix U ; by Lemma 3, PUQ^\dagger/p is an integral unitary transformation A of f into f . Then $U = P^\dagger AQ/p$. Conversely, if $U \equiv P^\dagger AQ/p$ has all its elements integral, where A is any integral automorph of f , then $PU = AQ$, $fQ = fAQ = fPU \sim fP$.

The general integral automorph of $[a, b, c]$ (assumed primitive) is

$$(17) \quad A = \begin{pmatrix} \frac{1}{2}(t - bu) & -cu \\ au & \frac{1}{2}(t + bu) \end{pmatrix},$$

where (t, u) runs over all integral solutions of (6). Forming $P_p^\dagger AP_\kappa$ ($\kappa < p$) we find only one element not termwise divisible by p ,

$$(18) \quad \Delta = au\kappa + \frac{1}{2}(t + bu).$$

To complete the proof of (7) it now suffices to prove

LEMMA 4. If fP_p is primitive, i. e. $p \nmid a$, there are exactly $\sigma - 1$ values κ ($0 \leq \kappa < p$) such that $p \mid \Delta$ by choice of t, u .

For by Lemma 2 it will follow for any primitive fP_κ , that there are precisely $\sigma - 1$ others equivalent to it in (13), by applying Lemma 4 to fV , where

$$V = \begin{pmatrix} p - \kappa & \kappa - p + 1 \\ -1 & 1 \end{pmatrix}, \text{ whence } fVP_p = fP_\kappa U \sim fP_\kappa, \quad U \equiv \begin{pmatrix} 1 & 1 - p \\ -1 & p \end{pmatrix}.$$

If $u \equiv 0 \pmod{p}$, $t \equiv \pm 2$ if $p > 2$, and

$$(19) \quad \frac{1}{2}(t + bu) \equiv \frac{1}{4}(t^2 - b^2u^2) = 1 - acu^2 \equiv 1, \quad \text{if } p = 2,$$

whence $\Delta \equiv \pm 1 \not\equiv 0$; disposing of all cases $\sigma = 1$ in Lemma 4.

If $d = -4$, we have also $t = 0$, $u = \pm 1$. If $f = [1, 0, 1]$, $\Delta = \pm \kappa$, $fP_p \sim fP_0$, and $\sigma = 2$ is effective in Lemma 4.

If $d = -3$, $f = [1, 1, 1]$, $t = \pm 1$, $u = \pm fP_p \sim fP_0 \sim fP_{p-1}$, $\sigma = 3$ is effective.

Finally consider d positive but not square, $\sigma >$

$$(20) \quad \Delta_k = au_k\kappa + \frac{1}{2}(t_k + bu_k),$$

whence

$$(21) \quad u_i\Delta_k - u_k\Delta_i = u_{i-k}, \quad t_i\Delta_k - t_k\Delta_i =$$

Now $p \mid u_i$ if and only if $\sigma \mid i$. If $i \equiv k \not\equiv 0 \pmod{\sigma}$ and by (21₁), $\Delta_i \equiv \pm \Delta_k \pmod{p}$. Thus we need consider for each such k , $\Delta_k \equiv 0$ determines a unique $\Delta_i \equiv \Delta_k \equiv 0$ for the same κ , $p \mid u_{i-k}$ by (21₁), $\sigma \mid$

THEOREM 3. Set $\tau = \{p - (d \mid p)\}/\sigma$. *Exact in Theorem 2 to each class C.*

3. With these developments freshly in mind

LEMMA 5. Let $p > 2$, $p \nmid a$, $d = b^2 - 4ac$, $b/\sigma = \sigma_{p,d} > 1$. Not counting the residue $\kappa \equiv -\frac{1}{2}b/a$ of residues $\kappa \pmod{p}$ for which $\Delta_k \equiv \pm 1 \pmod{p}$ but is

$$\sigma - 1 \text{ if } p \mid d, 2(\sigma - 1) \text{ if } p$$

i) Case $p > 2$, $p \mid d$. By (6), $t_i \equiv \pm 1$, $\kappa \equiv -\frac{1}{2}b/a$ appears from $\Delta_i \equiv 1$ if $t_i \equiv 2$, $\Delta_i \equiv$ remaining $\sigma - 1$ residues κ determined by $\Delta_i \equiv$ are distinct from each other and from $-\frac{1}{2}b/a$: for and $\sigma \nmid i - k$, (21₂) shows that Δ_k and $-\Delta_i$ are in

ii) Case $p > 2$, $p \nmid d$. If $\Delta_i \equiv \pm 1$ and $\Delta_k \equiv$ then by (21₁), $u_{i-k} \equiv \pm u_i \pm u_k$, which is impossible [For example, if $i + k$ is even, $i = h + l$, $k = u_{i-k} = u_l t_1$, etc. If $i + k$ is odd, transpose a term, way.]

THEOREM 4. Let v denote the number of primitive discriminant $p^2 d$ derivable from a primitive ambiguous d , by integral transformations of determinant p .

$$(22) \quad v = 1 \text{ if } p > 2, p \nmid d; v = 2 \text{ if } p >$$

If $p = 2$ let C contain the form $[a, b, c]$ in which a

$$(23) \quad \begin{aligned} \nu &= 0 \text{ if } d \equiv 0, 12, \text{ or } 28 \pmod{32}, \sigma = 1, \text{ and } b/a \text{ is odd;} \\ \nu &= 2 \text{ if } b/a \text{ is even and } \sigma = 1; \\ \nu &= 1 \text{ if } d \text{ is odd, or } \sigma > 1. \end{aligned}$$

We may assume $a \mid b$ in f . For the theorem to hold, (13) should contain $\nu\sigma$ primitive fP_κ improperly equivalent to themselves. If $fP_\kappa = fP_\kappa V$, where V is an integral matrix of determinant -1 , and B_1 is the matrix

$$\begin{pmatrix} 1 & b/a \\ 0 & -1 \end{pmatrix},$$

then $fB_1 = f$, $fP_\kappa = fB_1P_\kappa V$. By Lemma 3, if fP_κ is primitive, $B_1P_\kappa VP_\kappa^\dagger/p$ is an automorph (17); hence $P_\kappa VP_\kappa^\dagger/p$ is an improper integral automorph B of f , and $V = P_\kappa^\dagger BP_\kappa/p$. Conversely, if $V = P_\kappa^\dagger BP_\kappa/p$ is integral for some B , $fP_\kappa V = fBP_\kappa = fP_\kappa$, $fP_\kappa \sim fP_\kappa$.

The general improper integral automorph B of a primitive form $[a, b, c]$ in which $a \mid b$ is AB_1 , with A as in (17). Replacing u by $-u$, we find

$$(24) \quad B = \begin{pmatrix} \frac{1}{2}(t + bu) & \frac{1}{2}b(t + bu)/a - cu \\ -au & -\frac{1}{2}(t + bu) \end{pmatrix}.$$

The only element of $P_\kappa^\dagger BP_\kappa$ not obviously divisible by p is

$$(25) \quad \Gamma = au\kappa^2 + (t + bu)\kappa + \frac{1}{2}(t + bu)(b/a) - cu.$$

I. Case $p > 2$, $p \mid a$. Then $p \mid b$, $p \nmid c$, and by (14) fP_κ is primitive ($\kappa < p$). Now $p \mid d$, $p \nmid t$ by (6), $p \mid \Gamma$ if and only if

$$\kappa + \frac{1}{2}b/a \equiv cu/t \pmod{p}.$$

As t, u range over solutions of (6), u/t has exactly σ residues, $\nu = 1$.

II. Case $p > 2$, $p \nmid a$. Then fP_p is primitive and ambiguous. The unique residue $\kappa \equiv -\frac{1}{2}b/a$ appears from $\Gamma \equiv 0$ when $t \equiv \pm 2$, $u \equiv 0$; then $a\kappa^2 + b\kappa + c \equiv -\frac{1}{4}d/a$, and by (14), fP_κ is primitive if $p \nmid d$, imprimitive if $p \mid d$. In the remaining cases, $p \nmid au$ and

$$(26) \quad au\Gamma = (au\kappa + \frac{1}{2}(t + bu) + 1)(au\kappa + \frac{1}{2}(t + bu) - 1),$$

so that $\Gamma \equiv 0$ if and only if $\Delta \equiv \pm 1$. The residue $\kappa \equiv -\frac{1}{2}b/a$ cannot reappear, since it implies $4(\Delta^2 - 1) \equiv t^2 - 4 = du^2$. For d positive but not square (22) follows from Lemma 5. Our discussion is now complete also if $d < -4$ or d is a positive square. If $d = -4$, $f = [1, 0, 1]$, $\Delta = \kappa$, $fP_\kappa \sim fP_\kappa$ ($\kappa = 0, 1, p-1, p$), and since $\sigma = 2$, $\nu = 2$. If $d = -3$,

$f = [1, 1, 1]$, $\Delta = \kappa + 1$ or κ ; if now $p > 3$, fP_κ ($p-1$, p , $\frac{1}{2}(p-1)$), and since $\sigma = 3$, $\nu = 2$; if effective, fP_1 being imprimitive, $\nu = 1$.⁸

III. Case $p = 2$, $\sigma = 1$ (every u even). Then $\Gamma \equiv b/a$. Hence $\nu = 0$ if b/a is odd and a even, $\nu = (fP_2$ primitive). If b/a is even, the two primitive equivalent ($\sigma = 1$), $\nu = 2$.

IV. Case $p = 2$, $\sigma > 1$. Since $du^2 + 4$ is a square

$$(27) \quad d \equiv 5 \pmod{8}, 12 \pmod{16}, \text{ or } 0 \pmod{16}$$

In these cases, $\sigma = 2 - (d \mid 2) \geq \nu\sigma$. Hence $\nu = \nu > 0$ since fP_2 is primitive; if a is even, d is even, forms (13) are equivalent, $\Gamma \equiv u + b/a \pmod{2}$ is even

Remark. If $\gamma(d)$ denotes the number of primitive d , and d has exactly t distinct prime factors,

$$(28) \quad \gamma(d) = \begin{cases} 2^{t-2} & \text{if } d \equiv 4 \pmod{16}, \\ 2^{t-1} & \text{if } d \equiv 1 \pmod{4}, 8 \text{ or } 12 \pmod{16}, \\ 2^t & \text{if } d \equiv 0 \pmod{32}.^9 \end{cases}$$

Further the number of primitive ambiguous classes is $\gamma(d)$. By (28),

$$(29) \quad \begin{aligned} \gamma(4d) &= \gamma(d) \text{ if } d \equiv 1 \pmod{4}, 12 \pmod{16}, \text{ or } 0 \pmod{16}, \\ \gamma(4d) &= 2\gamma(d) \text{ if } d \equiv 4 \text{ or } 8 \pmod{16}, \text{ or } 0 \pmod{16}. \end{aligned}$$

Comparison of these facts with (23) yields (8).

For example, consider $d = -84$. Reduced forms

$$f_0 = [1, 0, 21] \text{ and } f_1 = [3, 0, 7]; f_2 = [5, 4, 5]$$

From these are derived the following of discriminant

$$\begin{aligned} f_0 &\rightarrow [1, 0, 84], [4, 0, 21]; f_1 \rightarrow [3, 0, 21], \\ f_2 &\rightarrow [5, \pm 2, 17]; f_3 \rightarrow [8, \pm 4, 5] \end{aligned}$$

The least positive non-square $d \equiv 0, 12$, or $28 \pmod{32}$; $d = 156$; the chains of reduced forms are $(1, 12, -3$

⁸ And in fact, $[1, 1, 7]$ is the only primitive reduced form for $d = -84$.

⁹ If d is a square double these values for $\gamma(d)$.

(2, 10, —7, 4, 5, 6, —6, 6, 5, 4, —7, 10, 2; and its negative); by (8), these four classes each contain ambiguous forms of only one type.

4. The term *m-derived classes of C* will signify the primitive classes derived from *C* by integral transformations of determinant *m*.

For $f = [a, b, c]$ and $f' = [a, -b, c]$ it is plain from (13) that $fP_\kappa \sim f'P_{p-\kappa}$ ($0 < \kappa < p$), $fP_0 \sim f'P_0$, $fP_p \sim f'P_p$. Hence by Lemmas 2 and 3 we have:

- (a) if *K* is a *p*-derived class of an ambiguous *C*, so is K^{-1} ;
- (b) the *p*-derived classes of a non-ambiguous *C* are non-ambiguous, and are the reciprocals of the *p*-derived classes of C^{-1} .

Both (a) and (b) extend at once to *m*-derived classes.

The *m*-derived classes of *C* are the same as the $(-m)$ -derived classes of C^{-1} .

THEOREM 5. *Theorem 2 holds with p replaced by any non-zero integer m.*

5. If in (13), fP_p is primitive and p^2n is represented therein, evidently *n* is represented in *f*; similarly for fP_κ . By induction from m^2d to $(pm)^2d$: an integral b. q. f. represents every number represented by any of its primitive derived classes [Gauss, D. A., Art. 166].

Hence, if a primitive f_1 of discriminant m^2d represents a prime *q* (or $-q$ or ± 1) represented by *f*, then f_1 is a $(\pm m)$ -derived form of *f*. For the property $f(q) > 0$ determines the class of *f* or its reciprocal.

THEOREM 6. *Let f and f_1 be primitive and of respective discriminants d and p^2d , any prime. Then f_1 is a $(\pm p)$ -derived form of f if and only if f_1 represents p^2q , where $\pm q$ is any prime (or 1) represented by f, and if and only if (4) holds.¹⁰*

6. Now (9) is an immediate consequence of (4) and Theorem 3.

7. Since *f* represents all integers represented in f_1 , each *p*-derived form f_1 of *f* has all the generic characters of *f*. An additional character

$$(30) \quad (f_1 | p), (-1 | f_1), (2 | f_1)$$

of f_1 occurs only in the following respective cases:

$$(31) \quad p > 2, p \nmid d; \quad p = 2, d \equiv 4, 8 \pmod{16}; \quad p = 2, d \equiv 16 \pmod{32}.$$

¹⁰ Pall I, 331, II, 494.

LEMMA 6. In the cases (31), among the $p - (d | p)$ primitive forms (13) occur equally many having either value ± 1 for (30).

We may suppose a prime to p .

i) $p = 2$. The forms (13) are now

$$(32) \quad f_0 = [4a, 2b, c], f_1 = [4a, 4a + 2b, a + b + c], f_2 = [a, 2b, 4c].$$

We find that if $d \equiv 4$ or $8 \pmod{16}$, then $f_2 \equiv a \pmod{4}$, and the primitive one of f_0, f_1 is $\equiv -a \pmod{4}$; if $d \equiv 16 \pmod{32}$, then $f_2 \equiv a \pmod{8}$ and the primitive one of f_0, f_1 is $\equiv 5a \pmod{8}$.

ii) $p > 2$. Now $\lambda = 2a\kappa + b$ is with κ an independent integral variable \pmod{p} . If $f_1 = fP_p$, $(f_1 | p) = (a | p)$; and for the rest,

$$(f_1 | p) = (a | p)(\lambda^2 - d | p).$$

Hence Lemma 6 is a consequence of

LEMMA 7. Let m be an integer, p an odd prime, $p \nmid m$. Then among the numbers

$$\lambda^2 - m \quad (\lambda = 0, 1, \dots, p-1),$$

there is one more quadratic non-residue than residue \pmod{p} .¹¹

Consider cases (31). Each primitive genus Γ of discriminant d has two p -derived genera Γ_i ($i = 1, -1$) differing only in the value i of (30). Let $\Gamma(n)$ denote the number of sets of representations of n by a representative system of classes of Γ . By (4) and Theorem 3

$$(33) \quad \Gamma_1(p^2n) = \Gamma_{-1}(p^2n) = \frac{1}{2}\{p - (d | p)\}\Gamma(n).$$

By (3), if m is prime to p , $\Gamma_i(pm) = 0$ ($i = 1, -1$); while

$$(34) \quad \begin{aligned} \Gamma_i(m) &= \Gamma(m), \quad \Gamma_{-i}(m) = 0, \\ [i &= (m | p), (-1 | m), (2 | m) \text{ according to (31)}]. \end{aligned}$$

There remain the cases (cf. (29))

$$(35) \quad p > 2, p \nmid d; \quad p = 2, d \equiv 1 \pmod{4}, 12 \pmod{16}, 0 \pmod{32}.$$

The p -derived classes of Γ form a single genus Γ' , for which, m and n being integers, $p \nmid m$,

¹¹ Thus there are $\theta = \frac{1}{2}\{p - (m | p)\}$ non-residues, $\theta - 1$ residues. This curious result must have appeared frequently in related forms.

$$(36) \quad \Gamma'(p^2n) = \{p - (d | p)\} \Gamma(n), \quad \Gamma'(pm) = 0, \quad \Gamma'(m) = \Gamma(m).$$

By induction from (33) and (36₁) we have

THEOREM 7. *An integer n has equally many sets of representations in every primitive genus of discriminant d which represents n .*

To determine these genera we proceed as follows. To begin with, if n' has no prime factors p such that d'/p^2 is a discriminant, then $r(n', d') = 0$ unless

$$(37) \quad p^a | n', \quad p^{a+1} \nmid n', \quad a \text{ odd imply } (d' | p) = 0 \text{ or } 1;$$

and then n' is represented in an unique genus Γ of discriminant d' .¹²

Generally, let k^2 be the largest square factor of n for which d/k^2 is a discriminant, and set $d = k^2 d'$, $n = k^2 n'$. Then n is represented only in the k -derived genera of the unique genus (if any) of discriminant d' which represents n' . The number of these genera, (37) being assumed, is 2^i , where i is the number of odd primes p such that $p | k$, $p \nmid d'$; this number being increased by 1 if

$$d' \equiv 1 \pmod{4}, \quad k \equiv 4 \pmod{8}, \quad \text{or } d' \equiv 4 \pmod{16}, \quad k \equiv 2 \pmod{4}; \quad \text{or}$$

$$d' \equiv 8, 16, 24 \pmod{32}, \quad k \text{ even, or } d' \equiv 12 \pmod{16}, \quad k \equiv 0 \pmod{4};$$

but increased by 2 if $d' \equiv 1 \pmod{4}$, $k \equiv 0 \pmod{8}$, or $d' \equiv 4 \pmod{16}$, $k \equiv 0 \pmod{4}$.

McGILL UNIVERSITY.

REFERENCES.

- G. Pall, I. *Mathematische Zeitschrift*, vol. 36 (1933), pp. 321-343; II. *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 491-509; III. *Bulletin of the American Mathematical Society*, vol. 41 (1935), pp. 373-4.
 Dirichlet, *Vorlesungen über Zahlentheorie*, 4th Ed.
 Lipschitz, *Journal für Mathematik*, vol. 53 (1857), pp. 238-259.
 G. B. Mathews, *Theory of Numbers* (1892).
 H. J. S. Smith, *Collected Mathematical Papers*, vol. I.

¹² Cf. Pall I, 337, footnote. It is easy to define this genus by the generic characters either of n' or of the ambiguous forms representing the prime factors common to n' and d' .

ABSTRACT LINEAR DEPENDENCE AND LATTICES.

By GARRETT BIRKHOFF.

1. *Introduction.* In a preceding paper,¹ Hassler Whitney has shown that it is difficult to distinguish theoretically between the properties of linear dependence of ordinary vectors, and those of elements of a considerably wider class of systems, which he has called "matroids."

Now it is obviously impossible to incorporate all of the heterogeneous abstract systems which are constantly being invented, into a body of systematic theory, until they have been classified into two or three main species. The purpose of this note is to correlate matroids with abstract systems of a very common type,² which I have called "lattices."

As this correlation is purely formal, the discussion will be descriptive rather than detailed.

2. *The fundamental construction.* Let us refer to Whitney's definition of a matroid as a set M whose subsets have a numerical rank function with certain properties $(R_1)-(R_3)$. It is easy to define mutual dependence by abstraction from the theory of vectors, and to see that if we suppress elements of rank zero, and regard mutually dependent elements as merely repeated occurrences of the same element, then we get a matroid M^* with the same structural properties as M . Moreover M^* has the additional property that no element is dependent on any other element.³

Now let us call a subset of M^* "linearly complete" if and only if it contains all elements dependent on it. By what we have just shown, distinct elements of M^* are distinct linearly complete subsets. Moreover

LEMMA 1. *The product $N \cdot N'$ of any two linearly complete subsets N and N' of M^* , is linearly complete.*

For suppose e any element of M^* not in $N \cdot N'$; by symmetry, we can suppose e not in N . Therefore

¹ H. Whitney, "On the abstract properties of linear dependence," *American Journal of Mathematics*, vol. 57 (1935), pp. 509-533. His main definition is in his section two.

² Cf. the author's "On the combination of subalgebras," *Proceedings of the Cambridge Philological Society*, vol. 39 (1933), pp. 441-464. This article will be referred to subsequently as "Subalgebras."

³ This assertion corresponds to S. MacLane's assumption "without loss of generality" of his conditions $(R_4)-(R_5)$ in his note on "Some interpretations of abstract linear dependence, etc.," *supra*.

$$\begin{aligned} 1 &= r(N + e) - r(N) && \text{[by definition of } N\text{]} \\ &\leq r(N \cdot N' + e) - r(N \cdot N') && \text{[by Whitney's Lemma 4].} \end{aligned}$$

Hence $r(N \cdot N' + e) \geq r(N \cdot N') + 1$, and e is not dependent on $N \cdot N'$, proving the lemma.

Observe now that M^* is itself linearly complete, and has the property $M^* \cdot N = N$ for any subset N . Finally, the operation of intersection is idempotent, commutative, and associative. Consequently

THEOREM 1. *The linearly complete subsets of M^* can be regarded as the elements of a new system $L(M^*)$, satisfying*

- (L_1) Any a of $L(M^*)$ and b of $L(M^*)$ determine a unique "product" $a \cap b$ in $L(M^*)$.
- (L_2) $a \cap a = a$, $a \cap b = b \cap a$, and $a \cap (b \cap c) = (a \cap b) \cap c$.
- (L_3) There exists a "unit" i [namely, M^*] such that $a \cap i = a$ irrespective of a .
- (L_4) In any sequence of products $a_1, a_1 \cap a_2, a_1 \cap a_2 \cap a_3, \dots$ some two terms are equal. [Remark: This is a simple corollary of the fact that $L(M^*)$ is finite].

Moreover distinct single elements of M^* appear as distinct single elements of $L(M^*)$.

3. *An incidental theorem.* We shall now prove a result which is obvious in the case in hand—provided we define the "join" $N \cup N'$ of N and N' directly as the product of all linearly complete subsets containing N and N' . (This definition is equivalent to defining $N \cup N'$ as the set of all elements linearly dependent on $N + N'$). The formal proof *in abstracto* has however been included, to give a neat new approach to the theory of abstract lattices.⁴

THEOREM 2. *It is an abstract consequence of Theorem 1 that any elements a and b of $L(M^*)$ determine a unique "join" $a \cup b$ with the properties*

⁴ It is extremely useful in applications. For instance, it shows that if we adjoin a purely formal "unity" i , we can subsume the theory of "effective implication" as defined by E. V. Huntington in his note "Effective equality and effective implication in formal logic," *Proceedings of the USA Academy of Sciences*, May, 1935, under the theory of lattices in the finite [but not the infinite!] case. Viewed in this light, his propositions 32, 33, 34, 36, 37 appear as extensions to the infinite case of Theorems 4.2, 6.1, 6.2, 6.3 of "Subalgebras."

With regard to precedence, it was Professor Huntington's paper which suggested Theorem 2 to me originally.

$$(L_5) \quad a \cap (a \cup b) = a \cup (a \cap b) = a.$$

$$(L_6) \quad a \cup a = a, \quad a \cup b = b \cup a, \quad \text{and} \quad a \cup (b \cup c) = (a \cup b) \cup c.$$

The main thing is to define $a \cup b$. To do this, note first that $i \cap a = a$ and $i \cap b = b$. Second, note that if $c \cap a = a$ and $c' \cap a = a$, then

$$(c \cap c') \cap a = c \cap (c' \cap a) = c \cap a = a.$$

Therefore if we set $c_0 = i$, and try to find successive elements c_1, c_2, c_3, \dots such that $c_k \cap a = a$, $c_k \cap b = b$, and yet $c_0 \cap \dots \cap c_k$ differs from every $c_0 \cap \dots \cap c_j$ [$j < k$], then by (L_4) we obtain after a finite number of attempts, an element $d = c_0 \cap \dots \cap c_n$ such that (1) $d \cap a = a$ and $d \cap b = b$ (2) if $c^* \cap a = a$ and $c^* \cap b = b$, then $c^* \cap d = c_1 \cap \dots \cap c_k$ for some $k < n$ —which, since

$$c^* \cap d = c^* \cap (d \cap d) = (c^* \cap d) \cap d = c_1 \cap \dots \cap c_k \cap d = d$$

[by iterated use of (L_2)], means $c^* \cap d = d$.

Conditions (1)-(2) and (L_2) show that d is determined uniquely by a and b ; we shall define $a \cup b$ as d . To prove that $a \cup b$ satisfies (L_5) - (L_6) is now merely a question of substituting from (1)-(2), and using condition (L_2) .

This completes the proof, the details of which the reader should be able to fill in without difficulty.

4. *Principal results.* But by definition, an "abstract lattice" is any system which satisfies (L_1) - (L_2) and (L_5) - (L_6) . Consequently, we have

THEOREM 3. $L(M^*)$ is an abstract lattice.

Not all lattices correspond to matroids; in particular, every lattice corresponding to a matroid contains only a finite number of elements. The facts as to which lattices correspond to matroids can however be summarized as follows.

A finite lattice corresponds to a matroid if and only if (1) if b and c are smallest distinct elements larger than a , then $b \cup c$ is a smallest element larger than b . (This is the "dual" [under the inversion of inclusion⁵] of the abstract property of composition subgroups used in the classic proof of the Theorem of Jordan-Hölder) (2) every element can be expressed as the "join" of elements of rank one.

⁵ It being observed that the definition of a lattice is invariant under inversion of meet and join. Property (1) amounts in the case in hand to the property that if two linearly complete subsets of M^* of rank $(n+1)$ intersect in a subset of rank n , then the rank of their join is $(n+2)$.

Moreover the "rank" $r(N)$ of any subset N of elements e_1, \dots, e_k of M^* is equal to the length of any chain [in the sense of the Theorem of Jordan-Hölder] in $L(M^*)$, between the lattice-element corresponding to the null subset, and the "join" of the lattice-elements corresponding (under Theorem 1) to the e_i . That is,

THEOREM 4. *The structure of $L(M^*)$ determines that of M^* —and hence that of M .*

5. *Theoretical consequences.* The results of the last sections suggest trying to interpret theorems on matroids as theorems on lattices, and conversely. This can be done in at least two cases.

The less interesting of these is Whitney's Theorem 3, which may be regarded as a reappearance⁶ of the first formula of Theorem 9.2 of "Subalgebras."

A more interesting correspondence concerns Whitney's Theorem 15 on separability, which is a generalization of a well-known theorem on graphs. This theorem may, in the light of the above results, be regarded as a specialization of a theorem⁷ on the "strong" uniqueness of the representation of any lattice as the direct product of irreducible (= non-separable) components.

6. *Geometrical correspondence.* In another paper,⁸ S. MacLane has shown that M^* also corresponds to a schematic configuration of dimensions $r(M^*) - 1$, and has pointed out a relationship with projective geometries. This defines an obvious correspondence *via* matroids between "matroid lattices" (or lattices satisfying conditions (1)-(2) of § 4) and schematic configurations.

In terms of this correspondence, results proved elsewhere⁹ by the author directly imply

THEOREM 5. *A schematic configuration is the direct product of a finite number of projective geometries and single points, if and only if the dual [under the inversion of join and meet] of the corresponding matroid lattice, is again a matroid lattice.*

⁶ This formula very likely has a long previous history.

⁷ Proved in the author's "On the lattice theory of ideals," *Bulletin of the American Mathematical Society*, vol. 40 (1934), p. 616, Theorem 2.

⁸ Cf. footnote 3, above. A much weaker result of the same sort was obtained by the author, in Theorem 5 of "On the structure of abstract algebras," to appear shortly in the *Proceedings of the Cambridge Philological Society*.

⁹ "Subalgebras," Theorem 10.2, and "Combinatorial relations in projective geometries," *Annals of Mathematics*, vol. 36 (1935), pp. 743-748.

(By the "direct product" of two projective geometries P and P^* is meant the system whose elements are the couples (A, A^*) [A any m -plane of P and A^* any m^* -plane of P^*], and in which (A, A^*) lies in (B, B^*) if and only if A lies in B and A^* in B^* . This is the inverse of the notion of separation into non-separable components mentioned above).

Theorem 5 and well-known results¹⁰ show that such schematic configurations can in general be realized by vectors in spaces whose coördinates lie in suitable finite fields.

SOCIETY OF FELLOWS,
HARVARD UNIVERSITY.

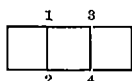
¹⁰ O. Veblen and W. H. Bussey, "Finite projective geometries," *American Transactions*, vol. 7 (1906), pp. 241-259.

SOME UNIQUE SEPARATION THEOREMS FOR GRAPHS.¹

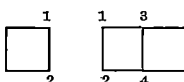
By SAUNDERS MACLANE.²

An abstract graph³ is a combinatorial structure which consists of a finite set of points or *vertices*, together with a number of edges. Each edge *joins* exactly two of the vertices. Such graphs are considered in topology, in the study of electrical circuits⁴ and in combinatorial analysis.⁵ Since the structure of such a graph becomes involved if the number of vertices is large, it is desirable to have methods of uniquely breaking up a graph into smaller parts or components. One such method is that of separating the graph at a single vertex. This is possible when the removal of the vertex in question disconnects the graph. The uniqueness of this process has been established by Whitney.⁶

Similar to the separation at a single vertex is the separation along a chain of connected vertices. Consider for example the graph



The vertices 1 and 2 together form a chain. If we drop these two vertices and all edges involving them, the graph falls into two connected pieces. If the separating vertices 1 and 2 be added to each piece, we have the two components



The second component may be further separated by the chain (3, 4). This gives three components, each one a square. Had we instead separated by the chain (3, 4) first, the result would have been the same.

¹ Presented to the Society, September 7, 1934.

² Sterling Research Fellow, Yale University.

³ M. A. Sainte-Laguë, "Les Réseaux," *Mémoires des Sciences Mathématiques*, Fasc. XVIII.

⁴ R. M. Foster, "Geometrical Circuits of Electrical Networks," *Bell Telephone System Technical Publications*, Mon. B-653.

⁵ E. Netto, *Lehrbuch der Combinatorik*, 2nd ed., pp. 292-308.

⁶ H. Whitney, "Non-separable and Planar Graphs," *Transactions of the American Mathematical Society*, vol. 34 (1932), p. 339.

This uniqueness property can be suitably extended: if any graph is separated by chains until the resulting components cannot be further separated, then the result is independent of the chains used in this separation. Two provisos are needed: 1. all the chains used in the separation must be geodesics (§ 3); 2. isomorphic components are considered as essentially identical (§ 4).

The plan of the proof is as follows: it is sufficient to consider any two geodesic chains B and C , and to construct a separation S' beginning with a separation by B , and a separation T' beginning with C , in such a way that S' and T' have the same results. First we use the geodesic property to show that B and C form together a symmetric figure. We then construct the desired separation S' as an alternating separation; first separation by B , then by C , then by B , and so on. A slight modification is made to insure that each separation is a separation by a single connected chain. The so modified separation S' then has the same result as the separation by a similarly constructed T' . This is established by showing that each stage of the alternating separation S' gives a finer subdivision of the original graph than does the preceding stage of the separation T' .

The separation process has two other basic properties: it yields a representation of the original graph (§ 1) and it assists in the choice of complete independent sets of cycles (§ 2).

1. *The Definition of Separation.* The graphs as defined in the introduction will be restricted by requiring that each edge join two distinct vertices and that no pair of vertices be joined by more than one edge. We can then conveniently represent the vertices by arabic numerals, while each edge may be denoted by the pair of vertices which it joins. Two vertices joined by an edge are called *contiguous*. Any symmetric binary relation $R(x, y)$ in which x and y range over a finite class (the class of vertices) can be interpreted as a graph by taking R to be the relation of contiguity.

Among the *subgraphs* of a graph G we consider only those subgraphs H such that two vertices of H contiguous in G are also contiguous in H . This restriction is natural in the relational interpretation. Such a subgraph $H \subseteq G$ is completely determined by the set of its vertices.

If a graph G and two subgraphs H_1 and H_2 are given, then $H_1 + H_2$ denotes the subgraph determined by the set of all vertices present in either H_1 or H_2 , while $H_1 - H_2$ denotes the subgraph determined by the set of vertices present in H_1 but not in H_2 .

A graph with $n > 2$ distinct vertices $1, 2, \dots, n$ and with the edges

$$(1\ 2), (2\ 3), \dots, (n-1\ n), (n, 1),$$

is called a *circuit*. Similarly, a *chain* is a graph of the type

$$(1\ 2), (2\ 3), \dots, (n-1, n)$$

with $n > 0$ distinct vertices. This chain has the *length* n , and is said to *join* its *end* vertices 1 and n . Its vertices are ordered, in that a vertex j on the chain is *between* the vertices i and k if $i < j < k$ or $k < j < i$.

The separation of a connected graph G by a connected⁷ subgraph H is an operation which yields a set of components, each of which is a subgraph of G . This set we denote by $H \times G$. To obtain it, dissect the graph $G - H$ into its connected pieces F_1, F_2, \dots, F_m . The set of subgraphs

$$H_1 = F_1 + H, H_2 = F_2 + H, \dots, H_m = F_m + H$$

is then the set $H \times G$. These components can also be characterized as follows:

THEOREM 1.1. *If H is a connected subgraph of a connected graph G , then every vertex of H belongs to every component of $H \times G$. Every vertex of $G - H$ belongs to just one component, and two such vertices belong to the same component if and only if they are joined in G by a chain containing no vertices of H .*

Any graph G is represented by the set of its components, in the following sense:

THEOREM 1.2. *If H is a connected subgraph of a connected graph G , then every vertex and edge of G is present in at least one component of $H \times G$.*

A set of components $H_1 \times G$ can be further separated by a subgraph H_2 of G if each component of $H_1 \times G$ which contains H_2 is separated by H_2 , while the other components of $H_1 \times G$ are left unchanged. The result of this separation we denote by $H_2 \times H_1 \times G$. The simplest uniqueness theorem⁸ then is

THEOREM 1.3. *If H_1 and H_2 are connected subgraphs of a connected graph G , and if H_1 and H_2 have no vertices in common, then*

$$H_1 \times H_2 \times G = H_2 \times H_1 \times G.$$

Proof. Denote the connected pieces of $G - (H_1 + H_2)$ by

⁷ This assumption here is essential; without it any graph could be separated into trivial components.

⁸ This theorem is essentially a generalization of Whitney's results for separation at single vertices.

$$G_1, G_2, \dots, G_n.$$

A vertex b of $H_1 + H_2$ will be called a *boundary vertex* of a piece G_i if and only if b is contiguous to a vertex of G_i . Two applications of Theorem 1.1 then show that the operation $H_1 \times H_2 \times G$ yields a component

$$(1) \quad H_1 + H_2 + \sum' G_i,$$

the sum being taken over all these pieces G_i which have boundary vertices both in H_1 and in H_2 , together with several components

$$(2) \quad H_1 + G_j,$$

one for each G_j whose boundary vertices are all in H_1 , and several components

$$(3) \quad H_2 + G_k,$$

one for each G_k whose boundary vertices are all in H_2 . Since these components, (1), (2), and (3), are symmetric in H_1 and H_2 , the uniqueness is established.

This theorem enables us to define the separation of a connected graph G by a disconnected subgraph H . For if H consists of the connected pieces H_1, H_2, \dots, H_m , we set

$$H \times G = H_1 \times H_2 \times \dots \times H_m \times G.$$

2. *Circuits and Separation.* If we consider a graph G as a one-dimensional topological complex, then it is known that in this complex we can choose a *complete set* of linearly independent one-dimensional circuits.⁹ Separation simplifies this choice, for the linearly independent circuits can all be chosen from the components *after* a separation $H \times G$.

To establish this result for a connected subgraph H , we agree to call a circuit C an *H-circuit* if $C - H$ is connected (in particular, C may be contained in H). An *H-circuit* C may contain no vertices of H whatsoever. If this is not the case, the vertices of C not in H form a chain with two ends a and b . The remaining vertices of C (i. e., those in H) form, together with a and b , a second chain with a and b as ends. Any *H-circuit* is thus contained in at least one component of $H \times G$.

LEMMA 2.1. *If a chain B in G contains no vertices of H , and if there is a circuit C containing B , then there is an H -circuit containing B .*

⁹ O. Veblen, *Analysis Situs*, 1st ed., p. 17.

Proof. If C is not itself an H -circuit, extend B along C until both end vertices are points of H . Then complete the desired circuit by joining these ends in H .

LEMMA 2.2. *Any circuit C in G may be represented as a sum (modulo 2) of H -circuits.*

Proof. Denote the connected pieces of $C - H$ by B_1, B_2, \dots, B_n . Each B_i is a chain, and by Lemma 2.1 is embedded in an H -circuit C_i , so that $C_i - H = B_i$. The sum

$$C + C_1 + \dots + C_n \text{ (modulo 2)}$$

has then no vertices not in H . It can thus be represented as a sum (modulo 2) of circuits in H ; that is, as a sum of H -circuits. These circuits, with the C_i , give the representation¹⁰ of C .

THEOREM 2.3. *If H is a connected subgraph of the connected graph G , then a complete set S of circuits in G can be constructed from $H \times G$ as follows: select in each component of $H \times G$ a complete set of circuits, combine all these sets to form a set T , and remove enough circuits from T to make the resulting set S linearly independent.*

Proof. The set is complete, for by Lemma 2.2 every circuit is a sum of H -circuits, while any H -circuit is contained in at least one component of $H \times G$. If H is a tree (i. e., contains no circuits), the set T itself can be shown to be linearly independent.

3. *Geodesic Separation.* The fundamental problem is to show that the result of separation by chains is essentially unique. This is not always true; for consider the graph which consists of three chains joining two vertices 1 and 2, with the edges

$$(1\,3), (3\,2), (1\,4), (4\,2), (1\,5), (5\,6), (6\,2).$$

If we separate this graph by the chain 1 3 2, we obtain a quadrilateral with vertices 1 2 3 4 and a pentagon with vertices 1 2 3 5 6. If instead we were to separate by the larger chain 1 5 6 2, we would obtain two pentagons, with vertices 1 2 3 5 6 and 1 2 4 5 6 respectively. The two results are essentially

¹⁰ A circuit C contained (in the usual sense) in a graph G need not be a subgraph of G according to our definition; such a C can however always be represented as a sum (modulo 2) of circuits which are subgraphs.

different.¹¹ This difference arises because the two separating chains 1 3 2 and 1 5 6 2 have the same ends, but different lengths.

To avoid the lack of uniqueness we can agree in such a case to separate only by the shorter of the two chains; that is, by a "geodesic" chain. In general, a chain C is *geodesic* in a graph G if every other chain contained in G and having the same end vertices as C is at least as long as C . Note that any subchain of a geodesic chain is geodesic.

A *complete geodesic separation* of a graph G will then proceed first by separating G by a geodesic chain B , then by separating each resulting component L by some chain B_1 , contained in and geodesic in L , and so on, until the final components are no longer separable in this way. The desired uniqueness theorem states that any two such complete geodesic separations S and T of the same graph G have essentially the same result. S will begin with a separation $B \times G$, while T starts with a separation $C \times G$. We first investigate the interrelations of the geodesics B and C .

This will involve the decomposition of a chain into subchains. A representation

$$D = D_1 + D_2 + \cdots + D_n$$

is said to be a *decomposition* of the chain D if each of the graphs D_i is a chain, if each vertex of D appears in one and only one of the chains D_i , and if the order of the sum on the right represents the order of the vertices of D . This last requirement means that if d_i , d_j , and d_k are vertices chosen from the three distinct chains D_i , D_j , and D_k , respectively, then d_j is between d_i and d_k in the chain D if and only if $i < j < k$ or $k < j < i$.

LEMMA 3.1. *Let B and C be geodesic chains in a graph G , and let f , g , and h be vertices belonging to both B and C . If g is between f and h in B , then g is also between f and h in C .*

Proof. Since g lies between f and h , there is a decomposition

$$B = B_1 + f + B_2 + g + B_3 + h + B_4,$$

in which f (likewise g and h) represents the subchain whose only vertex is f . If f were between g and h in C , then similarly

$$C = C_1 + g + C_2 + f + C_3 + h + C_4.$$

We thus have two distinct chains

¹¹ The results cannot be further reduced, for a circuit is not separable by any subgraph.

$$\begin{array}{c} f + B_2 + g + B_3 + h \\ f + C_3 + h \end{array}$$

joining the same vertices f and h . Both chains are geodesic, and hence have the same length. If the length of each chain B_i (or C_i) be denoted by the corresponding Greek letter, then

$$\beta_2 + 1 + \beta_3 = \gamma_3.$$

A similar argument for the chains joining g to h yields

$$\gamma_2 + 1 + \gamma_3 = \beta_3.$$

Substitution of the second equation in the first gives the contradiction

$$(\beta_2 + \gamma_2 + 2) + \gamma_3 = \gamma_3.$$

The intersection of two geodesic chains can now be described as follows:

THEOREM 3.2. *Let B and C be geodesic chains in a graph G , and let the intersection of B and C consist of exactly n connected pieces A_1, \dots, A_n . Then these A_k are chains, and there are two decompositions*

$$\begin{aligned} B &= B_1 + A_1 + B_2 + A_2 + \dots + B_n + A_n + B_{n+1}, \\ C &= C_1 + A_1 + C_2 + A_2 + \dots + C_n + A_n + C_{n+1}, \end{aligned}$$

where the B_i and C_i are chains, no two of which have a vertex in common. Each B_j and each C_j for $j = 2, \dots, n$ contains at least one vertex. The orientation of the A_k in both decompositions is the same; that is, we can denote the ends of each A_k by a_k and e_k in such a way that a_k is contiguous to a vertex of B_k and a vertex of C_k , while e_k is similarly contiguous to B_{k+1} and C_{k+1} . However, a_1 need not be contiguous to a vertex of B_1 , etc. Furthermore

$$\text{length } B_j = \text{length } C_j \quad (j = 2, \dots, n)$$

Finally, a vertex d on B_j ($j = 2, \dots, n$) is never contiguous to any vertex of C_i ($i \neq j$).

Proof. By suitably numbering the pieces A_k we can obtain the decomposition of B . A similar decomposition for C is possible, except that the chains A_k might appear here in a different order. But since each A_k contains at least one vertex, any such difference is impossible, by Lemma 3.1.

Furthermore, the decomposition of C might offer some A_k in an orientation different from that in the decomposition of B . Thus A_k might have ends

a_k and e_k , with a_k contiguous to B_k and C_{k+1} and e_k contiguous to B_{k+1} and C_k . We can assume without loss of generality that $k < n$, and we select a point d in A_n . We then have e_k between a_k and d in B , while a_k is between e_k and d in C , contrary to Lemma 3.1.

Next, the equality of the lengths follows because

$$e_{j-1} + B_j + a_j, \quad e_{j-1} + C_j + a_j$$

are two geodesic chains with the same ends. Finally, the non-contiguity of B_j and C_i is a consequence of the geodesic property of C .

The chains A_k ($k = 1, \dots, n$), B_i , and C_i ($i = 1, \dots, n+1$) into which B and C have thus been decomposed may be called the *atomic* chains of the figure. No two atomic chains have vertices in common.

4. *Isomorphic Components.* Two complete geodesic separations of a graph need not give identical results. Consider for instance the graph used in § 3. The separation by 132 gave a quadrilateral 1234 and a pentagon 1 2 3 5 6. The separation by the geodesic chain 1 4 2 would give the same quadrilateral, but a different pentagon 1 2 4 5 6. The two different pentagons are nevertheless still isomorphic.

In general, two graphs are said to be isomorphic¹² if their vertices can be put into a 1—1 correspondence which leaves invariant the relation of contiguity. It is important to note the effect of separation on isomorphic graphs.

THEOREM 4.1. *If two connected graphs G_1 and G_2 are isomorphic, and if G_1 be separated by a subgraph H_1 , while G_2 is separated by the corresponding subgraph H_2 , then there is a 1—1 correspondence between the components of $H_1 \times G_1$ and those of $H_2 \times G_2$, such that corresponding components are isomorphic.*

This theorem is a consequence of the general invariance of any logical property under an isomorphism.

5. *Alternating Separation.* Let a graph G and two geodesic chains B and C , with intersections as described in § 3, be given. We proceed to construct a separation S' , beginning with $B \times G$, and a separation T' , beginning with $C \times G$, such that S' and T' have the same results. S' is to be constructed in stages of alternating separations by B and by C , as indicated in

$$(1) \quad \dots C \times B \times C \times B \times G,$$

¹² Whitney, "On the Classification of Graphs," *American Journal of Mathematics*, vol. 55 (1933), p. 236.

and the first m stages of the separation will be denoted by $(C, B)^m \times G$. T' is to have a similar structure. However, the simple alternation indicated in (1) is not always sufficient.¹³ Thus it may happen that $B \times G$ yields a component L which does not contain all of G , such that the part of C in L consists of two separate chains D_1 and D_2 . Then instead of next separating L by C (or, more precisely, by the disconnected graph $D_1 + D_2$), it is desirable to separate L by a graph C' , obtained by adding to $D_1 + D_2$ enough vertices from B to insure that the result C' is a single chain. Similar modifications are to be made at each stage of the alternating separation.

This modified separation $(C, B)^m \times G$ will be defined by induction on m . At the same time we will establish the following properties of any component L of $(C, B)^m \times G$:

- 5.1. If one vertex of an atomic chain (see § 3, end) D belongs to L then D is contained in L ;
- 5.2. For every j , $j = 2, \dots, n$, either B_j or C_j belongs to L ;
- 5.3. For every k ; $k = 1, \dots, n$, A_k belongs to L .

For $m = 1$ we define

$$(C, B)^1 \times G = B \times G.$$

The three properties follow at once from Theorem 1.1.

Now let m be even and equal to $2p$, and let L be a component of the already defined separation $(C, B)^{2p-1} \times G$. L is to be separated by the chain C , after C has been modified as follows:

1. If C_1 is not present in L , drop C_1 from C ;
2. If C_{n+1} is not present in L , drop C_{n+1} from C ;
3. If C_j is not present in L ($j = 2, \dots, n$), replace C_j in C by the corresponding B_j .

These transformations of C yield a graph C_L , with the following properties:

- 5.4. C_L is a chain,
- 5.5. C_L is contained in L ,
- 5.6. C_L contains every vertex of C present in L ,
- 5.7. C_L is geodesic.

The first property is a consequence of the configuration of B and C , as described in Theorem 3.2. The construction of C_L , together with properties

¹³ Thus if G has nine vertices and the edges (12), (23), (34), (25), (54), (16), (67), (74), (18), (89), (94), and if B and C are the chains 1234 and 254 respectively, the simple alternating separations S' and T' will not give the same results.

5.1 and 5.2, gives the second and third properties. Finally, to show that C_L is geodesic, note that the first and second operations above transform C into a subchain, and thus leave it a geodesic, while, in virtue of Theorem 3.2, the third operation changes neither the length nor the ends of C , and so must leave it geodesic.¹⁴ C_L is thus geodesic in G , and therefore geodesic in L as well.

$(C, B)^{2p} \times G$ is now the set of all components obtained by separating each component L of $(C, B)^{2p-1} \times G$ by the corresponding chain C_L . The three properties 5.1, 5.2, and 5.3 of the separation are established by using Theorem 1.1 and noting that the separator C_L contains every A_k and just one of each pair B_j, C_j . A component M of $(C, B)^{2p} \times G$ which arises from a component L of $(C, B)^{2p-1} \times G$ by the separation $C_L \times L$ is said to have L as a *parent*. Every such component M has at least¹⁵ one parent.

The inductive definition of the alternating separation may be completed for the case of odd m by a treatment similar to that for even m , except that the letters B and C are to be everywhere interchanged. Thus each component M of $(C, B)^{2p} \times G$ is to be separated by a suitable modification B_M of the chain B . In any component L at any stage of separation $(C, B)^m \times G$ both the separators B_L and C_L can be constructed, although only one is to be used.

The phenomenon of isomorphic but not identical results discussed in § 4 can affect but a small part of a component L ; that part of L unaffected in this way we will call the *body* of L . Specifically, the body of L consists of all vertices of L , except that B_j is to be omitted from the body if C_j is not present in L , while C_j is to be omitted if B_j is not present ($j = 2, \dots, n$). We will denote the body of L by $\beta(L)$. A vertex not in $\beta(L)$ belongs to both B_L and C_L .

The "finer separation theorem" to be discussed in the next section applies only¹⁶ to those components L with $B_L \neq L$ and $C_L \neq L$. Such components we call *substantial*. A substantial component L contains at least one vertex not in B_L and one vertex not in C_L . These vertices are useful in establishing uniqueness properties.

6. Substantial Components in Alternating Separation. The ultimate

¹⁴ C would not necessarily remain a geodesic were C_1 to be replaced by B_1 . For this reason, the end chains B_1 and B_{n+1} cannot be treated in the same way as the intermediate chains B_2, \dots, B_n .

¹⁵ Two or more parents for one component can occur; see § 7.

¹⁶ For example, if G has six vertices and the edges (12), (23), (34), (25), (54), (16), (26), while $B = 254$ and $C = 1234$, then $(B, C)^2 \times G$ and $(C, B)^2 \times G$ differ in that the former has a non-substantial component 1254 not present in the latter.

agreement of the two separations $(C, B)^m \times G$ and $(B, C)^m \times G$ rests on the increasing fineness of the separation. Thus $C \times B \times G$, since it involves separations by both C and B , should subdivide G more finely than $C \times G$ alone. This "finer separation theorem" is

THEOREM 6.1. *If L is a substantial component of $(C, B)^m \times G$ ($m > 0$), then there exists a component M of $(B, C)^{m-1} \times G$, such that*

1. $\beta(L) \subset \beta(M)$,
2. M is substantial.

First note that the following conclusions could be added to those of the theorem.

3. If $g \in L$ and $g \notin C_L$ for some vertex g , then $g \in M$ and $g \notin C_M$.
4. The same for B_L and B_M .

These follow readily from the first conclusion. By the definition of the body of L , $g \in \beta(L)$. Hence $g \in \beta(M)$ and $g \in M$. Were $g \in C_M$, then either $g \in C$, whence by 5.6 $g \in C_L$, a contradiction, or $g \in B_j$ for some $j = 2, \dots, n$ such that C_j is absent from M . In the latter case $g \notin \beta(M)$, a contradiction.

The second conclusion of the theorem follows from the third conclusion, in virtue of the definition of substantial components.

The first conclusion itself may be proved by induction on m . If $m = 1$, we need only take $M = G$. If $m > 1$, we treat the typical case of even m . We select a parent L' of the given component L . The induction assumption yields a component M' in $(B, C)^{m-2} \times G$, with

$$\beta(L') \subset \beta(M').$$

The desired component M is now to be found as a component of $C_{M'} \times M'$ corresponding to L . Since L is substantial, there is a vertex g in L but not in C_L . But L has arisen by the separation $C_{L'} \times L'$, whence

$$(1) \quad C_L = C_{L'}.$$

Conclusion 3. above gives $g \in M'$, $g \notin C_{M'}$. Hence there is exactly one component M of $C_{M'} \times M'$ which contains g . This M , a component of $(B, C)^{m-1} \times G$, has the desired property 1, for if $h \in \beta(L)$, we prove $h \in \beta(M)$ by an argument in two cases, as follows.

Case 1. $h \notin C_L$. Then, by (1), $h \notin C_{L'}$. But since the vertex g chosen above to determine M lies together with h in the component L of $C_{L'} \times L'$.

there is by Theorem 1.1 some chain D in L' which has no vertices of C_L . By the induction assumption, D has no vertices of $C_{M'}$. Hence g and h must belong to M of $C_{M'} \times M'$. Finally, $h \in M$ and $h \notin C_{M'}$ imply $h \in$

Case 2. $h \in C_L$. It may be that h is in C_1 or in some other component of C_L . The desired conclusion $h \in \beta(M)$ is then immediate. Otherwise there is some j , ($j = 2, \dots, n$) such that $h \in C_j$, and C_j is contained in L . Hence both C_j and B_j are contained in L . Hence both C_j and B_j are contained in $\beta(M')$. Hence $C_j \subset M$ in the separation $C_{M'} \times M'$, while both B_j and C_j are in M , C_j and therefore h is in $\beta(M)$.

It is also necessary to show that the alternating sequence of separations comes to an end. This is the finiteness theorem.

THEOREM 6.2. *For a given graph G and fixed integer q such that, for all $m > q$,*

$$(B, C)^m \times G = (B, C)^q \times G$$

We thus introduce the notation for an "infinite" sequence of separations

$$(B, C)^\infty \times G = (B, C)^q \times G$$

Proof. Any component L in a separation (B, C) of G is separated from some parent component M by a separation such as E and $B_L = B_M$, so that L cannot be further separated from M . If now L should remain unchanged by the next separation F , it must remain unchanged by every subsequent separation B_L nor C_L separates L . That is, if L is separated from M at some stage of separation, then it must also be separated from M at every subsequent stage.

Return to the theorem and take q equal to the least integer such that the inequality

$$(B, C)^{m+1} \times G \neq (B, C)^m \times G$$

holds for some $m \geq q$. Then there is at least one component L_1 of G which is separated from one of its parents L_1 on the right. Derive L_2 from L_1 by L_2 , and of L_2 by L_3 , and so on, construct the sequence

$$L_0, L_1, L_2, \dots, L_q.$$

Since $L_0 = L$, the remarks of the preceding paragraph show that each L_i is different from its parent L_{i+1} ; in particular, each L_i has a smaller number of vertices than its parent. Since L_q cannot have more than q vertices, this shows that L_0 has no vertices at all, a contradiction.¹⁷

Any separation $H \times G$ is determinate, in that a vertex of G not in H is contained in exactly one component of $H \times G$. A useful consequence is

THEOREM 6.3. *If L and N are two substantial components of $(C, B)^m \times G$, such that*

$$\beta(L) \subset \beta(N);$$

then

$$L = N.$$

Proof. If L contains a vertex f contained in neither B nor C , then, by the determinateness mentioned above, L is the only component of $(C, B)^m \times G$ containing f . Since $f \in \beta(L)$, the result follows.

Suppose that no such vertex f is present. L consists then entirely of atomic chains from B and C . Since L is substantial, it contains either a pair of chains (B_j, C_j) for some $j = 2, \dots, n$, or else one of the pairs (B_1, C_1) , (B_1, C_{n+1}) , (B_{n+1}, C_1) , (B_{n+1}, C_{n+1}) . Now there is but one component of $(C, B)^m \times G$ which contains both B_j and C_j , as may be established from the determinateness property by induction. Since, by the definition of the body, both B_j and C_j belong to $\beta(L)$ and thus to $\beta(N)$ if they belong to L , the equality $L = N$ is again established. The other pairs above are treated in like manner.

The uniqueness theorem for substantial components now runs as follows:

THEOREM 6.4. *There is a 1 — 1 correspondence between the substantial components L of $(C, B)^\infty \times G$, and the substantial components M of $(B, C)^\infty \times G$, such that corresponding components are isomorphic.*

A suitable correspondence between the L 's and the M 's can be defined by the relation

$$(1) \quad \beta(L) = \beta(M).$$

This relation is equivalent to the relation

$$(2) \quad \beta(L) \subset \beta(M),$$

¹⁷ The value of q used here can be decreased considerably by a more elaborate analysis, based on the number n of chains common to B and C .

for, by the finer separation Theorem 6.1 there exists for each M an N in $(C, B)^\infty \times G$ with

$$\beta(M) \subset \beta(N).$$

This, combined with (2), shows that the conditions of Theorem 6.3 hold. The conclusion (1) follows at once. The converse, (1) implies (2), is immediate.

Furthermore, (1) is a 1—1 relation. The existence of an M for every L follows from the finer separation Theorem 6.1 and the equivalence of (2) and (1). The uniqueness of this M is a consequence of Theorem 6.3. The converse properties are established by similar proofs.

If L and M stand in the relation (1), then they are isomorphic. For (1), combined with 5.2 and the definition of a body of a component, shows that L and M differ, if at all, only in that, for certain values of j between 2 and n inclusive, L contains B_j but not C_j , while M contains C_j but not B_j (or vice versa). If in L we replace each such B_j by the corresponding C_j , or vice versa, we can transform L into M .

Each replacement $B_j \rightarrow C_j$ in this transformation is an isomorphism. Since B_j and C_j have the same number of vertices, we need only show that the contiguity relations of B_j and those of C_j are the same. In fact, both B_j and C_j are attached to the adjacent chains A_{j-1} and A_j in the same manner (see Theorem 3.2). No other contiguities between vertices of B_j and vertices of $L - B_j$ are possible. For suppose instead that a vertex g of B_j were so contiguous to a vertex h of L . This vertex is not in B , for B is a chain. If h were in C but not in B , then either $h \in C_j$, and both C_j and B_j belong to L , contrary to assumption, or $h \in C_i$ for $i \neq j$, contrary to Theorem 3.2. Hence h belongs to neither B nor C , and so $h \in \beta(L)$. By (2), $h \in \beta(M)$. Now in any separation $H \times G$ two contiguous vertices such as h and g must be present together in at least one component; in the separation $(B, C)^\infty \times G$ this must be the component M , for h is in only one component. But $g \in M$ contradicts the assumption that C_j but not B_j was contained in M . The invariance of the contiguity is thus established.¹⁸

7. *Duplicate and Non-substantial Components.* In a complete geodesic separation S a component may consist of just one edge e . However, the

¹⁸ The isomorphism of L to M is of a special type: any vertex g in L is identical with its corresponding vertex in M unless g is not contiguous to more than two points in L (alternatively, L is transformed into M by replacing certain "suspended" chains). The whole investigation could be formulated in terms of this narrower concept of isomorphism.

same edge e may be part of another component of S . In such a case we agree to drop the edge e from the set of components S . On the other hand, if e appears more than once as a component in S , but is not part of any larger component, then we agree to count e but once. This convention¹⁹ will not alter the fundamental Theorems 1.2 and 2.3.

More generally, any set W of components may contain a component L which is a "duplicate" because it is a subgraph of some other component in W . In a chain separation all such duplicate components are chains; hence the result of a complete geodesic separation is not altered if all duplicates are dropped as they arise.

The non-substantial components in a complete separation are either duplicates or arise from "bridges." By a bridge²⁰ e in a graph G we mean an edge which is contained in no circuit of G .

THEOREM 7.1. *If S is a complete geodesic separation of a connected graph G , then every bridge in G appears by itself as a component in S , and S contains no other components which are trees.*

Proof. A tree with more than one edge is always separable; hence the only trees in S are single edges of G . Any edge e of G which is not a bridge is contained in a circuit of G . For any separation $B \times G$, e is contained in a B -circuit (§ 2), and this circuit is contained in some component L of $B \times G$. Hence e belongs to a component which is not a tree. By induction, e is present in S in some component which is not a tree. Hence if e appears by itself in S , it is a duplicate.

Any bridge e' of G belongs to at least one component L in S and is a bridge in L . Were L to contain any edges other than e' , then L would be separable, contrary to the fact that S is a complete separation. Hence e' is the whole component L .

This theorem suggests that the treatment of duplicate components in a complete separation can be standardized as follows: Whenever a component L with a bridge e arises, then separate L successively by the two end-vertices of e . Then e appears by itself as a component. It is a duplicate unless it was a bridge in the original graph G .

8. Uniqueness Theorems for Chain Separation.

THEOREM 8.1. *Any two complete geodesic separations S and T of a connected graph G have the same result, in that there is a 1—1 correspond-*

¹⁹ Some such convention is necessary for the uniqueness theorems of § 8.

²⁰ E. Steinitz und H. Rademacher, *Vorlesungen über die Theorie der Polyeder*, p. 98.

ence between the components of S and the components of T such that corresponding components are isomorphic.

This theorem may be proved by induction on the number q of vertices in G . An enumeration of the cases shows it to be true for $q = 1, 2, 3, 4$. Assume then that it holds for all integers less than some q . Let the separations S and T begin with $B \times G$ and $C \times G$ respectively, and construct a complete separation S' beginning with $(C, B)^\infty \times G$. Also construct a complete separation T' beginning with $(B, C)^\infty \times G$ and continuing so that each substantial component of $(B, C)^\infty \times G$ is separated by T' in just the same way as the corresponding isomorphic component of $(C, B)^\infty \times G$ is separated by S' . By Theorem 6.4, $(C, B)^\infty \times G$ and $(B, C)^\infty \times G$ have the same result, except possibly for some components which are non-substantial; i. e., which are chains. Thence, by Theorems 4.1 and 7.1, S' and T' have the same result throughout. Now S and S' both begin with the separation $B \times G$. We can assume without loss of generality that $B \times G \neq G$; we can then apply the induction assumption to each component of $B \times G$ and obtain a 1—1 correspondence between components of S and S' which are not single edges. The correspondence can then be extended to include these single edges (Theorem 7.1). By a similar proof, T and T' have the same result. Hence S , through S' and T' , has the same result as does T .

A similar theorem can be obtained for a complete separation which uses geodesic chains whose lengths are all less than a fixed integer n_0 . This follows because all the separating chains used in the construction of $(B, C)^\infty \times G$ are at least as short as the longer of the two chains B and C .

For example, a chain of length 2 is always a geodesic. Hence if a graph is separated by single vertices and by edges (chains of length 2) as many times as possible, then the resulting set of components is independent of the particular vertices and edges used in the separation. In this case corresponding components are not only isomorphic, but also identical, for two chains B and C of length 2 cannot have more than one subchain in common (see Theorem 3.2 and the discussion of isomorphic components under Theorem 6.4).

The methods discussed in this paper for establishing the uniqueness of the separation of graphs by chains can be applied to study separation by circuits. The chief problem which arises is the development of a notion of "geodesic" circuits to exclude the cases in which separation is not unique.

GAUSSIAN DISTRIBUTIONS AND CONVERGENT INFINITE CONVOLUTIONS.

By AUREL WINTNER.

A result of Paley and Zygmund¹ [3] on Rademacher series suggests the following situation: *Every symmetric Bernoulli convolution is at the infinity at least as strongly damped as a Gaussian distribution function.* The object of the first part of the present note is the proof of a theorem to this effect. The result implies in particular that every symmetric Bernoulli convolution has *moments* of arbitrarily high order,

$$(1) \quad \mu_m(\phi) = \int_{-\infty}^{+\infty} x^m d\phi(x), \quad (m = 0, 1, 2, \dots),$$

and that these moments belong to a *determined* moment problem. As far as the asymptotic estimate by means of Gaussian distributions is concerned, it cannot be stated that the *density* of probability is majorized by a Gaussian density $C \exp(-\lambda x^2)$. In fact, there exist symmetric Bernoulli convolutions which are *nowhere* absolutely continuous (cf. [2], Example 6). It will, however, be shown that *the distribution function itself* is

$$(2) \quad O(\omega_\lambda(x)) \text{ as } x \rightarrow -\infty \text{ and } 1 + O(1 - \omega_\lambda(x)) \text{ as } x \rightarrow +\infty,$$

where

$$\omega_\lambda(x) = (\lambda/\pi)^{1/2} \int_{-\infty}^x \exp(-\lambda u^2) du$$

is the Gaussian distribution function with a sufficiently small dispersion parameter $\lambda > 0$. Since there need not exist a density, the situation is more delicate than in the case of "smooth" problems of the type of the distribution problem of the Riemann ζ -function (cf. [2], Theorem 16). It will turn out that *it is not necessary to restrict the considerations to the case of Bernoulli convolutions*. It may be emphasized that the Gaussian distribution function, majorizing the *convergent* infinite convolutions under consideration, belongs to a case of *divergence* of these infinite convolutions (cf. the end of the paper). The proof of (2) requires an adaptation to convolutions of an argument applied by Paley and Zygmund [3] to Rademacher series. The adaptation is made possible by the well-known algorithm of convolution moments (cf., e. g.,

¹ Numbers in brackets refer to the bibliography at the end of the paper.

[4]). This algorithm will play the rôle of the algorithm of orthogonal functions of Rademacher. The result (2) for convolutions is interesting also in view of a theorem of H. Steinhaus which has been shown in [5], that there exists for every $\epsilon > 0$ a function f such that its convolution the Fourier transform of which is

$$O(\exp(-|t|^{2-\epsilon}))$$

as $t \rightarrow \pm \infty$.

Let

$$(3) \quad \sigma_1, \sigma_2, \dots, \sigma_n, \dots$$

be a sequence of distribution functions which are, for instance, concentrated on a finite interval. On considering the Fourier transform ("Faltung")

$$\sigma_1 * \dots * \sigma_n$$

and using the notation (1), it is easily shown (cf. [4])

$$\mu_m(\sigma_1 * \dots * \sigma_n) = [\mu(\sigma_1) + \dots + \mu(\sigma_n)]^m$$

where $[]^m$ is a symbolical power affecting the μ with the usual rules of words,

$$(4) \quad \mu_m(\sigma_1 * \dots * \sigma_n) = \sum C_{ij\dots} \mu_i(\sigma_1) \mu_j(\sigma_2) \dots$$

where the summation runs through all collections (i, j, \dots) of non-negative integers i, j, \dots for which $i + j + \dots = m$, and $C_{ij\dots}$ is a multinomial coefficient

$$C_{ij\dots} = (i + j + \dots)! / i! j! \dots$$

It is easy to see (cf. [3]) that

$$(5) \quad C_{2i+2j\dots} \leq m^m C_{ij\dots}, \quad \text{if } i + j + \dots = m$$

Let, in particular,

$$(6) \quad \mu_{2m+1}(\sigma_1) = 0, \mu_{2m+1}(\sigma_2) = 0, \dots, \mu_{2m+1}(\sigma_n) = 0$$

where $m = 0, 1, 2, \dots$. Then from (4)

$$\mu_{2m+1}(\sigma_1 * \dots * \sigma_n) = 0,$$

while

$$\mu_{2m}(\sigma_1 * \dots * \sigma_n) = \sum C_{2i+2j\dots} \mu_{2i}(\sigma_1) \mu_{2j}(\sigma_2) \dots$$

where the summation runs through all collections (i, j, \dots) for which $2i + 2j + \dots = 2m$. Hence from (5)

$$(7) \quad \mu_{2m}(\sigma_1 * \dots * \sigma_n) \leq m^m \sum C_{ij} \dots \mu_{2i}(\sigma_1) \mu_{2j}(\sigma_2) \dots, \quad (m = 1, 2, \dots),$$

where the summation runs through all collections for which $i + j + \dots = m$.

Let

$$(8) \quad \sigma_n(x) = \beta(x/b_n), \quad (b_n > 0, n = 1, 2, \dots),$$

where $\beta(x)$ denotes the symmetric Bernoulli distribution function which is equal to 0, $\frac{1}{2}$ or 1 according as x lies on the left, in the interior or on the right of the interval $-1 \leq x \leq 1$. Then (6) is satisfied, while

$$\mu_{2m}(\sigma_n) = b_n^{2m} \quad (m = 0, 1, 2, \dots; n = 1, 2, \dots).$$

Hence from (7)

$$\mu_{2m}(\sigma_1 * \dots * \sigma_n) \leq m^m \sum C_{ij} \dots b_1^{2i} b_2^{2j} \dots,$$

where the summation runs through all collections of n non-negative integers i, j, \dots for which $i + j + \dots = m$ and so the factor \sum of m^m is the m -th power of $b_1^2 + b_2^2 + \dots + b_n^2$. Consequently,

$$(9) \quad \mu_{2m}(\sigma_1 * \dots * \sigma_n) < (Bm)^m, \quad (m, n = 1, 2, \dots)$$

where

$$B = \sum_{n=1}^{\infty} b_n^2.$$

It will be supposed that this series is convergent. This assumption is, according to [2], necessary and sufficient for the existence of a distribution function τ such that

$$(10) \quad \sigma_1 * \dots * \sigma_n \rightarrow \tau \text{ as } n \rightarrow +\infty,$$

i. e., for the convergence of the infinite convolution

$$(10a) \quad \tau = \sigma_1 * \sigma_2 * \dots,$$

where σ_n is defined by (8). Now it is easy to see that $\mu_{2m}(\tau)$ is finite for every m and that, furthermore,

$$(11) \quad (0 \leq) \mu_{2m}(\tau) \leq (Bm)^m, \quad (m = 1, 2, \dots).$$

In fact, it is clear from (9) and (1) that, for every $R > 0$,

$$\int_{-R}^R x^{2m} d\tau_n(x) < (Bm)^m, \text{ where } \tau_n = \sigma_1 * \dots * \sigma_n.$$

Since, from (10),

$$\int_{-R}^R x^{2m} d\tau_n(x) \rightarrow \int_{-R}^R x^{2m} d\tau(x), \quad n \rightarrow +\infty,$$

in virtue of Helly's theorem on term-by-term integration, it follows that

$$\int_{-R}^R x^{2m} d\tau(x) \leq (Bm)^m,$$

which implies (11) by letting $R \rightarrow +\infty$.

The estimate (11) shows that the series

$$\sum_{m=1}^{\infty} [\mu_{2m}(\tau)]^{-1/(2m)}$$

is divergent. Hence

$$\mu_m(\tau) = c_m \quad (m = 0, 1, 2, \dots),$$

is a determined moment problem in virtue of Carleman's classical criterion.

It is also seen from (11) and from the convergence of the series

$$\sum_{m=1}^{\infty} m!^{-1} \lambda^m B^m m^m, \text{ where } m!^{-1} m^m \sim (2\pi m)^{-1/2} e^m,$$

for sufficiently small values of $\lambda > 0$ that the series

$$\sum_{m=0}^{\infty} m!^{-1} \lambda^m \int_{-\infty}^{+\infty} x^{2m} d\tau(x) = \sum_{m=0}^{\infty} m!^{-1} \lambda^m \mu_{2m}(\tau)$$

is convergent, and so

$$\int_{-\infty}^{+\infty} \exp(\lambda x^2) d\tau(x) < +\infty$$

for every sufficiently small $\lambda > 0$. Thus for every $x > 0$

$$\text{const.} > \int_x^{+\infty} \exp(\lambda u^2) d\tau(u) \geq \exp(\lambda x^2) \int_x^{+\infty} d\tau(u) = \exp(\lambda x^2) [1 - \tau(x)],$$

or

$$1 - \tau(x) = O(\exp(-\lambda x^2)), \quad x \rightarrow +\infty,$$

while

$$\tau(x) = O(\exp(-\lambda x^2)), \quad x \rightarrow -\infty$$

by reasons of symmetry. This completes the proof of (2), since λ is arbitrary and

$$\begin{aligned} \exp((\lambda - \epsilon)x^2) \int_x^{+\infty} \exp(-\lambda u^2) du &< \int_x^{+\infty} \exp((\lambda - \epsilon)u^2 - \lambda u^2) du \\ &= \int_x^{+\infty} \exp(-\epsilon u^2) du \rightarrow 0, \quad 0 < x \rightarrow +\infty; \end{aligned}$$

hence

$$\int_x^{+\infty} \exp(-\lambda u^2) du = O(\exp(-(\lambda - \epsilon)x^2)), \quad x \rightarrow +\infty,$$

for every fixed positive $\epsilon (< \lambda)$.

The above proof applies also in cases where (10a) is not a Bernoulli convolution. In fact, it is sufficient to assume that (10a) is a convergent infinite convolution of arbitrary symmetric distribution functions σ_n for which

$$B < +\infty, \quad \text{where} \quad B = \sum_{n=1}^{\infty} \mu_2(\sigma_n),$$

and

$$(12) \quad (\mu_2(\sigma_n))^m \geq \alpha^m \mu_{2m}(\sigma_n), \quad (m = 0, 1, 2, \dots; \quad n = 1, 2, \dots),$$

where α is a sufficiently small positive constant.

First, the symmetry of σ_n means that

$$\sigma_n(x) + \sigma_n(-x) = 1.$$

Since this implies (6), the inequality (7) is applicable. Hence, from (5) and (12),

$$\mu_{2m}(\sigma_1 * \dots * \sigma_n) \leq m^m \alpha^{-m} \left(\sum_{j=1}^n \mu_2(\sigma_j) \right)^m,$$

and so

$$(9a) \quad \mu_{2m}(\sigma_1 * \dots * \sigma_n) \leq (Am)^m, \quad (m, n = 1, 2, \dots),$$

where

$$A = \alpha^{-1} B \geq \alpha^{-1} \sum_{j=1}^n \mu_2(\sigma_j)$$

is independent of m and n . The balance of the proof is the same as above.

A sufficient condition for (12) is that

$$(8a) \quad \sigma_n(x) = \rho(x/b_n), \quad (b_n > 0, \quad n = 1, 2, \dots),$$

where $\rho(x)$ is an arbitrary symmetric distribution function the spectrum of which is a bounded set and contains at least two points. In fact, the latter conditions imply that

$$|\mu_m(\rho)|^{1/m} < \text{const. and } \mu_2(\rho) \neq 0.$$

On the other hand,

$$\mu_m(\sigma_n) = b_n^m \mu_m(\rho)$$

by the definition of σ_n . Thus (12) is satisfied. The convergence of

$$B = \sum_{n=1}^{\infty} b_n^2$$

is necessary and sufficient for the convergence of the infinite convolution (10a) also when $\rho \neq \beta$. On placing

$$x = y\sqrt{B_n}, \text{ where } B_n = b_1^2 + \cdots + b_n^2,$$

the finite convolution

$$\sigma_1(x) * \cdots * \sigma_n(x)$$

tends, as $n \rightarrow +\infty$, to a Gaussian distribution function $\omega = \omega(y)$ whenever $B = +\infty$ and $b_n < \text{const.}$, while it tends to the distribution function $\tau(y\sqrt{B})$ whenever $B < +\infty$, i. e., whenever (10a) is a convergent infinite convolution.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] G. H. Hardy, "A theorem concerning Fourier transforms," *Journal of the London Mathematical Society*, vol. 8 (1933), pp. 227-231.

[2] B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88, more particularly § 6.

[3] R. E. A. C. Paley and A. Zygmund, "On some series of functions (1)," *Proceedings of the Cambridge Philosophical Society*, vol. 26 (1930), pp. 337-357, more particularly Lemma 2. Also A. Zygmund, *Trigonometrical series* (1935), p. 124.

[4] A. Wintner, "On an application of diophantine approximations to the repartition problems of dynamics," *Journal of the London Mathematical Society*, vol. 7 (1932), pp. 242-246.

[5] A. Wintner, "On symmetric Bernoulli convolutions," *Bulletin of the American Mathematical Society*, vol. 41 (1935), pp. 137-138.

ON CONVERGENT POISSON CONVOLUTIONS.

By AUREL WINTNER.

Let

$$\pi = \pi(x) = \pi(x; a, q), \quad -\infty < x < +\infty,$$

where

$$a > 0 \quad \text{and} \quad 0 < q < 1,$$

denote the one-dimensional distribution function which has at $x = 0$ and $x = a$ the jumps $1 - q$ and q respectively. Thus $\pi(x)$ is a step-function which is equal to

$$0, \quad 1 - q \quad \text{or} \quad 1,$$

according as x is in the interval

$$(-\infty, 0), \quad (0, a) \quad \text{or} \quad (a, +\infty).$$

Clearly

$$(1) \quad \mu_k(\pi) = a^k q, \quad \text{hence} \quad \mu_2(\pi) - [\mu_1(\pi)]^2 = a^2 q(1 - q),$$

and

$$(2) \quad L(t; \pi) = 1 - q(1 - e^{iat}),$$

where $\mu_k(\sigma)$ denotes the k -th moment

$$(1a) \quad \mu_k(\sigma) = \int_{-\infty}^{+\infty} x^k d\sigma(x), \quad (k = 1, 2, \dots; \mu_0(\sigma) = 1),$$

and $L(t; \sigma)$ the Fourier transform

$$(2a) \quad L(t; \sigma) = \int_{-\infty}^{+\infty} e^{itx} d\sigma(x), \quad (-\infty < t < +\infty),$$

of the distribution function σ . Poisson's law of seldom events deals, in its classical form ¹ [2], with the limit of a properly reduced convolution of a finite number of distribution functions π . The present note considers infinite convolutions of distinct distribution functions π without any reduction, that is to say infinite convolutions

$$(3) \quad \pi_1(x) * \pi_2(x) * \pi_3(x) * \dots * \pi_n(x) * \dots,$$

where

$$(4) \quad \pi_n(x) = \pi(x; a_n, q_n),$$

and the asterisk is the convolution operator.

¹ Numbers in brackets refer to the bibliography at the end of the paper.

It has been shown in [1] that if both series

$$(5) \quad \sum_{n=1}^{\infty} \mu_1(\sigma_n), \quad \sum_{n=1}^{\infty} \{\mu_2(\sigma_n) - [\mu_1(\sigma_n)]^2\}$$

are convergent, then so is the infinite convolution

$$(6) \quad \sigma_1(x) * \sigma_2(x) * \dots$$

Now the series (5) take in the case (3) the forms

$$(7) \quad \sum_{n=1}^{\infty} a_n q_n, \quad \sum_{n=1}^{\infty} a_n^2 q_n (1 - q_n)$$

in virtue of (4) and (1), and it is shown by examples like

$$(8) \quad q_n \sim n^{-4}, \quad a_n^2 q_n (1 - q_n) = 1 \quad (a_n q_n \sim n^{-2}, n \rightarrow \infty)$$

that the convergence of the first series (7) is compatible with the divergence of the second series (7). It is nevertheless true that the convergence of the first series (7) alone implies the convergence of (3). It is sufficient to prove that *the convergence of the first series (5) alone implies the convergence, even the absolute convergence, of (6), whenever no $\sigma_n(x)$ has a negative x in its spectrum.* Now in the latter case

$$\mu_1(\sigma_n) = \int_{-\infty}^{+\infty} |x| d\sigma_n(x),$$

and it has been shown in [8] that the assumption

$$\sum_{n=1}^{\infty} \int_{-\infty}^{+\infty} |x| d\sigma_n(x) < +\infty$$

always implies the absolute convergence of (6). Hence the situation is this:

The infinite Poisson convolution (3) is absolutely convergent whenever

$$(9) \quad \sum_{n=1}^{\infty} a_n q_n < +\infty.$$

On the other hand, (9) is not implied by the absolute convergence of (3); cf. (2) and (14b).

It will always be supposed that (9) is satisfied. The distribution function represented by (3) will be denoted by

$$(3a) \quad \rho = \rho(x).$$

Thus

$$(10) \quad L(t; \rho) = \prod_{n=1}^{\infty} L(t; \pi_n)$$

according to the general theory [1] of infinite convolutions. It is seen from (8) that (9) may be satisfied also when

$$(11) \quad a_n \rightarrow +\infty \quad \text{as} \quad n \rightarrow +\infty.$$

This does not contradict the fact, proven in [1], that the convergence of (3) implies the limit relation

$$(9a) \quad \pi_n \rightarrow \delta,$$

where δ denotes the distribution function

$$(12) \quad \delta(x) = \frac{1}{2}(1 + \operatorname{sgn} x).$$

In fact, the n -th term of the series (9) is the area of the rectangle which is bordered by the curves

$$y = \pi_n(x) \quad \text{and} \quad y = \delta(x),$$

and so (9a) is implied by the condition

$$(9b) \quad a_n q_n \rightarrow 0,$$

which is compatible with (11).

In view of Lévy's deduction [4] of the Gauss law, it is worth while to mention that, as a consequence of the convergence criterion (9), *there exist infinite convolutions such that*

$$(13) \quad \mu_1(\sigma_n) = 0 \quad \text{and} \quad \mu_2(\sigma_n) = 1 \quad \text{for every } n,$$

although (6) is convergent. If (6) is convergent, then the finite convolution

$$\sigma_1(x) * \sigma_2(x) * \cdots * \sigma_n(x),$$

when reduced to have the dispersion 1, does not tend to the Gaussian distribution function of dispersion 1 but to the distribution function (12), for which $\mu_2(\delta) = 0$. Now the Fourier transforms of the distribution functions

$$(13a) \quad \sigma(x) \quad \text{and} \quad \sigma(x + b)$$

are identical up to the factor e^{itb} . Hence, if

$$(13b) \quad \sum_{n=1}^{\infty} b_n$$

is convergent, the convergence of (6) is necessary and sufficient for the convergence of the "translated" infinite convolution

$$\sigma_1(x + b_1) * \sigma_2(x + b_2) * \cdots$$

Thus the existence of a convergent infinite convolution (6) satisfying (13) is, in view of (1), (9) and (13b), a consequence of the fact that (8) is compatible with (9), the difference of the first moments μ_1 of the distribution functions (13a) being equal to b in virtue of (1a).

The addition rule of spectra [1], when applied to (3), shows that the spectrum of (3) always consists of $x = 0$ and of the closure of the positive values

$$(14) \quad a_i + a_j + a_k + \cdots,$$

where i, j, k, \cdots is an arbitrary finite collection of distinct positive integers. Thus the spectrum of (3) is a bounded set if and only if

$$(14a) \quad \sum_{n=1}^{\infty} a_n < +\infty,$$

a condition implying (9). The smoothness criterion (21), to be proven later on, is compatible with (14a). On the other hand, (3) cannot be a continuous function if

$$(14b) \quad \sum_{n=1}^{\infty} q_n < +\infty.$$

In fact, it follows by a complete induction from the definition of a convolution as an integral, that the jump of the function (3) at $x = 0$ is never less than

$$\prod_{n=1}^{\infty} (1 - q_n),$$

and is therefore positive in the case (14b). Further, it is clear from the spectrum rule (14) that the spectrum of (3) is the whole interval $[0, +\infty]$, whenever

$$(14c) \quad \sum_{n=1}^{\infty} a_n = +\infty \text{ and } a_n \rightarrow 0.$$

This does not imply the continuity of (3), since (14c) is compatible with (14b) and (9). An example to this effect is

$$q_n = a_n^2 = (n+1)^{-2}.$$

Now if both (14b) and (14c) are satisfied, then the discontinuity points of

(3) lie dense in $[0, +\infty]$ in virtue of Theorem 35 of [1]. The spectrum of (3) is a sequence of points

$$(15) \quad 0, c_1, c_2, \dots, \text{ where } c_n \rightarrow +\infty,$$

if, for instance, (11) is satisfied or if a_n is independent of n , conditions necessitating (14b) in view of (9). On the other hand, not every distribution function which has a spectrum of the structure (15) may be represented in the form (3). An example to this effect is the Poisson distribution function [2], for which $c_n = n$ and the jump at $x = n$ (≥ 0) is $(n!e)^{-1}$.

An approach to the investigation of smoothness properties of $\rho(x)$ consists of obtaining, if possible, estimates of $L(t; \rho)$ for $t \rightarrow \pm\infty$; cf. [7]. In this direction it will first be shown that if

$$(16) \quad \sum_{n=1}^{\infty} q_n^2 < +\infty,$$

where

$$(16a) \quad q_n \leq \frac{1}{3},$$

and if (9) is satisfied, then

$$(17) \quad \log |L(t; \rho)| < -Ct^2 f(t) + A,$$

where C and A are positive constants and $f(t)$ denotes the even function

$$(18) \quad f(t) = \sum_{an|t| \leq 1} q_n a_n^2.$$

The inequality (17) is understood to hold for every t , with the proviso that $\log |L|$ denotes $-\infty$ if $L = 0$.

The proof of (17) depends on a combined adaptation of arguments used in [7] and [9]. First,

$$(19) \quad \log |L(t; \pi)| < -q(1 - \cos at) + 8q^2, \text{ if } q \leq \frac{1}{3}.$$

In fact,

$$\log |L(t; \pi)| + q(1 - \cos at)$$

is the real part of the function

$$\log L(t; \pi) + q(1 - e^{iat})$$

and so not larger than the absolute value of this function, hence, according to (2),

$$\leq \sum_{n=2}^{\infty} |q(1 - e^{iat})|^n / n \leq \sum_{n=2}^{\infty} (2q)^n / n < 8q^2,$$

if $q \leq \frac{1}{3}$. This proves the inequality (19), which is, according to (16a), applicable for every $q = q_n$. Now

$$1 - \cos t \geq Ct^2$$

for some $C > 0$ and for every t in the interval $-1 \leq t \leq 1$. Hence

$$1 - \cos a_n t \geq Ca_n^2 t^2$$

for every $a_n \leq 1/|t|$. Consequently,

$$\sum_{a_n|t| \leq 1} q_n (1 - \cos a_n t) \geq \sum_{a_n|t| \leq 1} q_n Ca_n^2 t^2$$

for every t . On combining this inequality with (18) and (19) and using the assumption (16), it is seen that

$$\sum_{a_n|t| \leq 1} \log |L(t; \pi_n)| < -Ct^2 f(t) + A$$

for some sufficiently large $A > 0$. This implies

$$\sum_{n=1}^{\infty} \log |L(t; \pi_n)| < -Ct^2 f(t) + A$$

in virtue of $|L(t; \pi_n)| \leq 1$ and therefore completes, according to (10), the proof of (17).

Since the spectrum of $\rho(x)$ lies at the right of the origin, $\rho(x) = 0$ for $x < 0$ but not for every x , so that $\rho(x)$ cannot be regular-analytic along the x -axis. On the other hand, it is possible to delimit a class of sequences (4) for which $\rho(x)$ has derivatives of arbitrarily high order for every x . According to [7], this will be the case whenever

$$(20) \quad L(t; \rho) = O(\exp -|t|^\gamma), \quad t \rightarrow \pm \infty,$$

for some $\gamma > 0$. Now, on using (17), it is easy to prove that if

$$(21) \quad \frac{1}{2} < \nu < 1 < \lambda + \nu$$

and if (4) is such that

$$(22) \quad n^{-\delta-\lambda} < a_n < n^{\delta-\lambda} \quad \text{and} \quad n^{-\delta-\nu} < q_n < n^{\delta-\nu}$$

for every $\delta > 0$ and every $n > N = N_\delta$, then $\rho(x)$ has for every x derivatives of arbitrarily high order, (20) being satisfied for every

$$(23) \quad \gamma < (1 - \nu)/\lambda.$$

At the first sight, the restrictions (21) seem to be artificial. It turns out, however, that *the inequalities (21) are about the best of their kind.*

First, (9) and (16) are satisfied in virtue of (21) and (22). Also, (16a) is implied by (16) for large n , and since in (23) the sign of equality is anyway excluded, no generality is lost by supposing that (16a) holds not only for large n but for every n . Thus (17) is applicable. Now from (21)

$$2\lambda + \nu > 1;$$

hence, as $|t| \rightarrow \infty$,

$$(24) \quad \sum_{n-\lambda|t| \leq 1} n^{-2\lambda-\nu} \sim \int_{|t|^{1/\lambda}}^{\infty} u^{-2\lambda-\nu} du = B |t|^{(1-2\lambda-\nu)/\lambda},$$

where $B > 0$. Since δ in (22) is arbitrarily small, it is seen from (21), (22) and (24) by straight-forward estimates that

$$\sum_{an|t| \leq 1} a_n^2 q_n > |t|^{(1-2\lambda-\epsilon)/\lambda}$$

for every $\epsilon > 0$ and for every $|t| > T = T_\epsilon$. Hence, from (18) and (17),

$$\log |L(t; \rho)| < -|t|^{(1-\nu-\epsilon)/\lambda}$$

for every $\epsilon > 0$ and for every $|t| > T' = T'_\epsilon$. This means that (20) is true for every γ satisfying (23), q. e. d.

That the inequalities (21) are, in the main, as good as possible, is seen as follows. First, $\nu > 1$ is impossible for any λ , since $\nu > 1$ implies (14b), hence a discontinuity for $\rho(x)$. On the other hand, on letting

$$2\nu \rightarrow 1 + 0 \quad \text{and} \quad 2\lambda \rightarrow 1 + 0,$$

which is compatible with (21), it is seen that

$$(1 - \nu)/\lambda \rightarrow 1 - 0,$$

while (20) holds for every γ satisfying (23). Hence (21) permits in (20) every $\gamma < 1$. Now (20) cannot hold in the limiting case $\gamma = 1$. In fact, (20) with $\gamma = 1$ implies, as shown in [7], the analyticity of $\rho(x)$ along the x -axis, when, as a fact, $\rho(x)$ cannot be analytic along the x -axis. This proves that $\nu < \frac{1}{2}$ is impossible for some λ . The assumption $1 < \lambda + \nu$ has been made in view of (9). Finally, it was necessary to exclude in (21) the equality signs, since one had to take care of the small exponents $\pm \delta$ in (22).

These considerations hold, of course, also under the restriction $\lambda = \nu$, in which case (21) goes over into $\frac{1}{2} < \nu < 1$ and the rectangle of small area, mentioned in connection with (12), is nearly a square.

Without using the language of the convolution theory, Schoenberg has effectively shown [5] that the distribution problem of the sequence

$$(25a) \quad \log 1 - \log \phi(1), \log 2 - \log \phi(2), \dots, \log n - \log \phi(n), \dots$$

depends on the particular case

$$(25) \quad q_n = 1/p_n \sim 1/(n \log n), \quad a_n = -\log(1 - 1/p_n) \sim q_n$$

of the infinite convolutions treated in the present note, $\phi(n)$ being Euler's function and p_n the n -th prime number. Schoenberg obtains (10) from a direct consideration of Schur and not from the theory of convolutions. An application of the theory of convolutions to arithmetical functions which are more general than (25a) will be given by Schoenberg in a paper referred to in [1]. It is seen from (14c) that in the case (25) the spectrum of $\rho(x)$ is the whole interval $[0, +\infty]$. On the other hand, Schoenberg [5] has proven in an interesting manner that $\rho(x)$ has, in the case (25), no discontinuity points. It is not known whether Schoenberg's $\rho(x)$ is or is not absolutely continuous; the criterion (20) fails, since (25) implies $\nu = \lambda = 1$. Correspondingly, Schoenberg's proof for the continuity of $\rho(x)$ depends on arithmetical properties of his a_n , more particularly on their linear independence, and not merely on questions (22) of rough order of magnitude.

It is interesting that the existence of a rough order of magnitude implies a high degree of smoothness for the symmetric case of infinite "Bernoulli convolutions" also. The result of [7] is precisely to this effect but imposes a strong restriction of regularity on the increase of the correction factor. The method used above admits of a simpler and general treatment if one replaces (17) by the corresponding inequality given in [9].

In order to prove this, let $\beta = \beta(x)$ be the distribution function which has both at $x = -1$ and $x = 1$ the jump $\frac{1}{2}$, i. e., the distribution function for which $L(t; \beta) = \cos t$, and let the sequence $\{b_n\}$ be such that

$$(26) \quad \sum_{n=1}^{\infty} b_n^2 < +\infty, \quad \text{where } b_n > 0.$$

Then the infinite Bernoulli convolution

$$(27) \quad \sigma(x) = \beta(x/b_1) * \beta(x/b_2) * \dots$$

is convergent and has the Fourier transform

$$(28) \quad L(t; \sigma) = \prod_{n=1}^{\infty} \cos(b_n t).$$

On placing

$$(29) \quad N(t) = \sum_{b_n | t| \geq 1} 1,$$

it is easy to see that

$$(30) \quad L(t; \sigma) = O(\exp\{-C[N(2t) - N(t)]\}) \quad (t \rightarrow \pm \infty)$$

for some $C > 0$. In fact, $N(t)$ clearly is the number of values n satisfying the inequality $b_n \leq 1/|t|$, where $b_n > 0$. Hence, if $t > 0$, there are at least

$$(29a) \quad N(2t) - N(t)$$

values of n satisfying both inequalities

$$(29b) \quad 1/(2t) \leq b_n \leq 1/t.$$

Now for all these (29a) values of n ,

$$(29c) \quad 0 < \cos(b_n t) \leq \cos \frac{1}{2}$$

in virtue of (29b). Hence it is seen from (28) that

$$(30a) \quad |L(t; \sigma)| \leq (\cos \frac{1}{2})^{N(2t) - N(t)}$$

for every $t > 0$. Since $L(t; \sigma)$ and $N(t)$ are even functions in view of (28) and (29), the estimate (30) is implied by (30a), the constant C being the logarithm of $1/\cos \frac{1}{2}$, and so $C > 0$. Now (30) implies the desired analogue to (22), and even more:

If (26) is satisfied and if

$$(31) \quad n^{\delta-\alpha} < b_n < n^{\delta-\alpha}$$

holds for a fixed $\alpha > 0$, for every $\delta > 0$ and for every $n > K = K_\delta$, then (27) has derivatives of arbitrarily high order for every x . Furthermore, (27) is or is not regular analytic along the x -axis according as $\alpha < 1$ or $\alpha > 1$. On the other hand, (27) may possess derivatives of arbitrarily high order for every x also in cases where the decrease of b_n is stronger than that allowed by (31), an example to this effect being

$$(32) \quad b_n = \exp(-n^\kappa) \quad (0 < \kappa < \frac{1}{2}).$$

Finally, in the case

$$(33) \quad b_n = c^n, \quad c = 2^{-1/h} \quad (h = 1, 2, \dots),$$

where the decrease of b_n is still stronger, (27) acquires derivatives of arbitrarily high order for every x as

$$c \rightarrow 1 - 0, \text{ i. e., } h \rightarrow +\infty,$$

for a fixed h there being assured the existence of a continuous $(h-1)$ -th derivative for every x .

The last result is interesting also because (27) is, according to [3], not an absolutely continuous function if $0 < c < \frac{1}{2}$ (for $c = \frac{1}{2}$ one obtains the classical Cantor function).

Consider first (31). Since (31) holds for every $\delta > 0$, it is easily seen from (29) and (31) that

$$|t|^{(1-\epsilon)/\alpha} < N(t) < |t|^{(1+\epsilon)/\alpha}$$

for every $\epsilon > 0$ and for every $|t| > T = T_\epsilon$. This clearly implies

$$N(2t) - N(t) > |t|^{(1-\eta)/\alpha}$$

for every $\eta > 0$ and for every $|t| > T' = T'_\eta$. It therefore follows from (30) that

$$L(t; \sigma) = O(\exp(-|t|^\gamma))$$

for every $\gamma < 1/\alpha$. Thus there always exists a $\gamma > 0$, and $\gamma = 1$ is admissible whenever $\alpha < 1$. Finally, if $\alpha > 1$, then

$$\sum_{n=1}^{\infty} b_n < +\infty,$$

i. e., the spectrum of (27) is a bounded set and so (27) cannot be regular analytic along the x -axis. The spectrum of the simplest case, $b_n = n^{-\alpha}$, has further been discussed in [3].

Consider next (32). The function $N(t)$ is, according to (29) and (32), the largest integer not exceeding $(\log |t|)^{1/\kappa}$, where $|t| > 1$. A straightforward reduction yields therefore for large values of $|t|$ the asymptotic formula

$$N(2t) - N(t) \sim (\kappa^{-1} \log 2) (\log |t|)^{(1-\kappa)/\kappa}.$$

Thus, since $0 < \kappa < \frac{1}{2}$,

$$N(2t) - N(t) > \epsilon^{-1} \log |t|$$

for every $\epsilon > 0$ and for every $|t| > T = T_\epsilon$. Hence from (30)

$$L(t; \sigma) = O(|t|^{-1/\epsilon})$$

for every fixed $\epsilon > 0$, an estimate implying the existence of all derivatives of (27) for every x . In fact, it is seen from the inversion formula [4],

$$\sigma(x) = \sigma(0) + \int_{-\infty}^{+\infty} (2\pi it)^{-1} (1 - e^{-itx}) L(t; \sigma) dt, \quad \left(\int_{-\infty}^{+\infty} = \lim \int_{-T}^T \right),$$

of the Fourier transform (2a) that, for a fixed $h > 0$, the estimate

$$(34) \quad L(t; \sigma) = O(|t|^{-h}) \quad (t \rightarrow \pm \infty)$$

is a sufficient condition for the existence of continuous derivatives $\sigma^{(k)}(x)$ of every order $k \leq h - 1$.

The remaining case (33) is now easily treated. In fact, from (28) and (33),

$$L(t; \sigma) = \prod_{n=1}^{\infty} \cos(2^{-n/h} t).$$

Hence

$$(35) \quad L(t; \sigma) = \prod_{j=0}^{h-1} \prod_{m=1}^{\infty} \cos(2^{-m} 2^{-j/h} t).$$

This may be written, in virtue of Vieta's identity

$$\prod_{m=1}^{\infty} \cos(2^{-m} z) = z^{-1} \sin z,$$

in the form

$$L(t; \sigma) = \prod_{j=0}^{h-1} (2^{-j/h} t)^{-1} \sin(2^{-j/h} t).$$

Consequently,

$$L(t; \sigma) = \prod_{j=0}^{h-1} O(|t|^{-1}) = O(|t|^{-h}),$$

which completes the proof in view of the criterion (34).

It may be mentioned that $\sigma(x)$ is, in the case (33), *symmetrically convex*, i. e., such that the derivative $\sigma'(x)$ is a non-increasing function of $|x|$. In fact, if two distribution functions are symmetrically convex, then so is their convolution. Now on denoting by $\sigma_h(x)$ the $\sigma(x)$ which belongs to (33), it is seen from (35) that

$$L(t; \sigma_h) = \prod_{j=0}^{h-1} L(2^{-j/h} t; \sigma_1),$$

i. e.,

$$\sigma_h(x) = \sigma_1(x) * \sigma_1(2^{1/h} x) * \dots * \sigma_1(2^{(h-1)/h} x).$$

Thus it is sufficient to show that $\sigma_1(x)$ is symmetrically convex. Now $L(t; \sigma_1)$ is Vieta's product and so

$$L(t; \sigma_1) = t^{-1} \sin t = \frac{1}{2} \int_{-1}^1 e^{itx} dx;$$

hence

$$\sigma'_1(x) = \frac{1}{2} \text{ if } |x| < 1 \text{ and } \sigma'_1(x) = 0 \text{ if } |x| > 1.$$

Thus $\sigma_1(x)$ is symmetrically convex.

Similarly, the $\sigma(x)$ belonging to $b_n = n^{-1}$ is symmetrically convex, since it may be written in the form

$$\sigma(x) = \sigma_1(x) * \sigma_1(3x) * \sigma_1(5x) * \dots$$

This $\sigma(x)$ is regular for every x , since in (31) the correction factors $n^{\pm\delta}$ are superfluous and so

$$L(t; \sigma) = O(\exp(-|t|^\gamma))$$

holds not only for every $\gamma < 1/\alpha = 1$ but for $\gamma = 1$ also.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

-
- [1] B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88, more particularly § 4.
 - [2] A. Kintchine, *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, Berlin, 1933, p. 17.
 - [3] R. Kershner and A. Wintner, "On infinite Bernoulli convolutions," *American Journal of Mathematics*, vol. 57 (1935), pp. 541-548. In formulae (24) and (25) of this paper the minus sign is missing in the exponents.
 - [4] P. Lévy, *Calcul des probabilités*, Paris, 1925.
 - [5] I. Schoenberg, "Ueber die asymptotische Verteilung reeller Zahlen modulo eins," *Mathematische Zeitschrift*, vol. 28 (1928), pp. 171-199, more particularly § 5.
 - [6] A. Wintner, "Ueber die statistische Unabhängigkeit der asymptotischen Verteilungsfunktion inkommensurabler Partialschwingungen, *ibid.*, vol. 36 (1933), pp. 618-629 and vol. 37 (1933), pp. 479-480.
 - [7] A. Wintner, "On analytic convolutions of Bernoulli distributions," *American Journal of Mathematics*, vol. 56 (1934), pp. 659-663.
 - [8] A. Wintner, "On the differentiation of infinite convolutions," *ibid.*, vol. 57 (1935), pp. 363-366.
 - [9] A. Wintner, "On symmetric Bernoulli convolutions," *Bulletin of the American Mathematical Society*, vol. 41 (1935), pp. 137-138.

A NOTE ON THE CONVERGENCE OF INFINITE CONVOLUTIONS.

By AUREL WINTNER.

It is known¹ that the infinite convolution $\sigma_1 * \sigma_2 * \dots$ of distribution functions $\sigma_n = \sigma_n(x)$ is convergent if and only if the product

$$(1) \quad \prod_{n=1}^{\infty} L(t; \sigma_n), \quad \text{where} \quad L(t; \sigma_n) = \int_{-\infty}^{+\infty} e^{itx} d\sigma_n(x),$$

is uniformly convergent in every fixed finite t -interval. This criterion for the convergence of $\sigma_1 * \sigma_2 * \dots$ has the disadvantage of involving a variable parameter t . Convergence criteria which contain only the first and second moments of every σ_n have been proven in the paper referred to before, where a general theory of $\sigma_1 * \sigma_2 * \dots$ also is developed. The object of the present note is to point out a simple convergence criterion which does not contain a parameter and applies also when the moments of second or first order do not exist. Let $\delta > 0$ and

$$(2) \quad \sum_{n=1}^{\infty} M_{\delta}(\sigma_n) < +\infty, \quad \text{where} \quad M_{\delta}(\sigma_n) = \int_{-\infty}^{+\infty} |x|^{\delta} d\sigma_n(x).$$

If (2) holds for $\delta = \delta_0$, it need not hold for $\delta < \delta_0$, although it requires the less of every σ_n the smaller is δ . Now if $0 < \delta \leq 1$, then (2) implies the absolute convergence of $\sigma_1 * \sigma_2 * \dots$. For if $0 < \delta \leq 1$, then $|\sin tx| \leq |tx|^{\delta}$. Now $|e^{itx} - 1| \leq 2|\sin(tx/2)|$. Hence $|e^{itx} - 1| \leq 2|tx|^{\delta}$. Consequently, since $L(0; \sigma_n) = 1$,

$$|L(t; \sigma_n) - 1| \leq \int_{-\infty}^{+\infty} 2|tx|^{\delta} d\sigma_n(x) = 2|t|^{\delta} M_{\delta}(\sigma_n).$$

Hence it is seen from (2) that the product (1) is absolutely and uniformly convergent in every fixed finite t -interval. Furthermore, (2) implies the convergence of $\sigma_1 * \sigma_2 * \dots$ only when $\delta \leq 1$. For let $\sigma_n(x) = 0$ or $\sigma_n(x) = 1$ according as $x \leq 0$ or $x \geq 2/n$, while $\sigma_n(x) = 1/2$ if $0 < x < 2/n$. Then $2M_{\delta}(\sigma_n) = (2/n)^{\delta}$, hence (2) holds for every $\delta > 1$, although

$$\prod_{n=1}^{\infty} L(t; \sigma_n) = \prod_{n=1}^{\infty} \exp(it/n) \cos(t/n)$$

is not convergent for every t .

THE JOHNS HOPKINS UNIVERSITY.

¹ B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta-function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88.

DETERMINATION OF A VAN DER CORPUT-LANDAU ABSOLUTE CONSTANT.

By RICHARD KERSHNER.

It has been shown by van der Corput and Landau¹ that there exists an absolute constant γ with the property that if $f(x)$ be a real-valued function possessing in the finite interval $[a, b]$ a second derivative which is nowhere less than a fixed positive constant r , then

$$(1) \quad \left| \int_a^b \cos f(x) dx \right| \leq \gamma r^{-1/2},$$

it being understood that γ is independent of $[a, b]$ as well as of $f(x)$. This result has several applications in the analytic theory of numbers. Let γ_0 denote the least permissible value γ . The best estimate of γ_0 so far known seems to be

$$\gamma_0 \leq 32^{1/2} = 5.657 \dots$$

We shall determine the actual value of γ_0 by verifying a suggestion of Wintner that the maximum of the expression on the left of (1) occurs when $f(x)$ is a parabola such that $f''(x) \equiv r$. The precise result is included in the following statement:

Let $f(x)$ be real-valued and possess a second derivative $f''(x) \geq r > 0$ in $[a, b]$. Then

$$\left| \int_a^b \cos f(x) dx \right| \leq \int_{-\beta_0}^{\beta_0} \cos(rx^2/2 + \mu_0) dx = \gamma_0 r^{-1/2},$$

where

$$\beta_0 = [(\pi - 2\mu_0)/r]^{1/2},$$

the constant² $\mu_0 = -0.725 \dots$ being determined as the only root of

¹ Cf. E. Landau, *Vorlesungen über Zahlentheorie*, vol. 2 (1927), p. 60; or E. Landau, *Einführung in die Differentialrechnung und Integralrechnung* (1934), p. 307; also for further references E. C. Titchmarsh, "On van der Corput's method and the zeta function of Riemann," *Quarterly Journal of Mathematics*, vol. 2 (1932), p. 173.

² The values of μ_0 and γ_0 given here were obtained by interpolation into the table of Fresnel integrals included in G. N. Watson, *A Treatise on the Theory of Bessel Functions* (1922), and are correct to $\pm .002$.

$$\int_0^{(\pi/2-\mu)^{1/2}} \sin(x^2 + \mu) dx = 0$$

in $-\pi/2 \leq \mu \leq \pi/2$, so that

$$\gamma_0 = 2\sqrt{2} \int_0^{(\pi/2-\mu_0)^{1/2}} \cos(x^2 + \mu_0) dx = 3.327 \dots$$

It will be sufficiently clear from the proof that the function $f(x) = rx^2/2 + \mu_0$ and the interval $[a, b] = [-\beta_0, \beta_0]$ lead, for any fixed r , to a maximum for the expression on the left of (1) which is unique up to the obviously permissible translations which do not affect the value of $\cos f(x)$.

The method of van der Corput and Landau consists, in the main, in introducing $f(x)$ as an independent variable and applying the second mean value theorem. In the present paper we shall work directly with elementary properties of the cosine curve. In fact, the proof follows directly from the following lemma.

LEMMA. *Let $f_1(x)$, $f_2(x)$ be real-valued and possess first derivatives $f'_1(x)$, $f'_2(x)$ in the intervals $[a, b_1]$, $[a, b_2]$ respectively.*

Suppose that

$$(i) \quad f_1(a) = f_2(a)$$

and denote this common value by μ . Let k be the integer for which

$$(k - 1/2)\pi \leq \mu < (k + 1/2)\pi$$

and suppose that

$$(ii) \quad f_1(b_1) = f_2(b_2) = (k + 1/2)\pi.$$

Let

$$(iii) \quad f'_1(x) > 0 \text{ in } a < x \leq b_1, \quad f'_2(x) > 0 \text{ in } a < x \leq b_2,$$

and let

$$(iv) \quad f'_1(a) \geq f'_2(a) \geq 0.$$

Finally, let

$$(v) \quad f'_1(x_1) \geq f'_2(x_2).$$

for every x_1 in $[a, b_1]$, where $x_2 = x_2(x_1)$ denotes the uniquely determined point of $[a, b_2]$ for which

$$f_1(x_1) = f_2(x_2).$$

[This function $x_2(x_1)$ exists and is unique in $[a, b_2]$ by virtue of (iii)]. Then

$$(2) \quad \left| \int_a^{b_1} \cos f_1(x) dx \right| \leq \left| \int_a^{b_2} \cos f_2(x) dx \right|.$$

Proof. We notice first that these conditions cannot all be satisfied unless $b_2 \geq b_1$. In fact, an application of the mean value theorem of the differential calculus to the inverse functions of $f_1(x)$ and $f_2(x)$ shows that the function $\phi(x_1)$ defined by

$$\phi(x_1) = x_2 - x_1 \equiv x_2(x_1) - x_1$$

is a monotone, non-decreasing function of x_1 in $[a, b_1]$ by virtue of condition (v). It is also clear from (iii)-(v) that

$$(3) \quad f_1(x) \geq f_2(x) \text{ in } [a, b_1].$$

In the sequel we shall suppose that $k=0$. This is permissible, since

$$\left| \int_a^b \cos f(x) dx \right| = \left| \int_a^b \cos \{f(x) - k\pi\} dx \right|$$

for any integer k .

We now consider the cases $\mu \geq 0$ and $\mu < 0$ separately.

If $\mu \geq 0$, then (2) is trivial, for $\cos x$ is non-negative and decreasing in $[0, \pi/2]$, so that (3) implies

$$\cos f_1(x) \leq \cos f_2(x) \text{ in } [a, b_1].$$

Consequently

$$\int_a^{b_2} \cos f_2(x) dx \geq \int_a^{b_1} \cos f_2(x) dx \geq \int_a^{b_1} \cos f_1(x) dx > 0.$$

If $\mu < 0$, then there exist, by virtue of (i)-(v), constants c_1, c_2, d , such that $a < c_1 < d < c_2 < b_2, d < b_1$, and

$$(4) \quad f_1(c_1) = f_2(c_2) = 0, \quad f_1(d) = -f_2(d).$$

Since $\cos x$ is increasing in $[-\pi/2, 0]$ and decreasing in $[0, \pi/2]$, it is clear from (3) and (4) that

$$(5) \quad \cos f_1(x) \geq \cos f_2(x) \text{ in } [a, d], \quad \cos f_1(x) \leq \cos f_2(x) \text{ in } [d, b_1].$$

Furthermore, from (i) and (4),

$$\cos f_1(a) = \cos f_2(a), \quad \cos f_1(d) = \cos f_2(d).$$

Hence, on denoting by $A_1 (> 0)$ the area of the region bounded by the two curves

$$y = \cos f_1(x) \text{ in } [a, d], \quad y = \cos f_2(x) \text{ in } [a, d],$$

and by $A_2 (> 0)$ the area of the region bounded by the three curves

$y = \cos f_1(x)$ in $[d, b_1]$, $y = \cos f_2(x)$ in $[d, b_2]$, $y = 0$ in $[b_1, b_2]$,

we clearly have

$$\int_a^{b_2} \cos f_2(x) dx - \int_a^{b_1} \cos f_1(x) dx = A_2 - A_1.$$

Consequently, if we call $A_3 (> 0)$ the area of the region bounded by the three curves

$y = \cos f_1(x)$ in $[c_1, d]$, $y = \cos f_2(x)$ in $[d, c_2]$, $y = 1$ in $[c_1, c_2]$,

then

$$(6) \quad \int_a^{b_2} \cos f_2(x) dx - \int_a^{b_1} \cos f_1(x) dx = (A_2 + A_3) - (A_1 + A_3).$$

But $A_1 + A_3$ is precisely the area of the region bounded by the three curves

$y = \cos f_1(x)$ in $[a, c_1]$, $y = \cos f_2(x)$ in $[a, c_2]$, $y = 1$ in $[c_1, c_2]$;

and $A_2 + A_3$ is the area of the region bounded by the four curves

$y = \cos f_1(x)$ in $[c_1, b_1]$, $y = \cos f_2(x)$ in $[c_2, b_2]$, $y = 1$ in $[c_1, c_2]$,
 $y = 0$ in $[b_1, b_2]$.

Since $\phi(x_1)$ is a monotone, non-decreasing function on $[a, b_1]$, it follows that

$$A_1 + A_3 \leq (c_2 - c_1)(1 - \cos \mu) \leq c_2 - c_1$$

and

$$A_2 + A_3 \geq c_2 - c_1.$$

Thus the expression on the right of (6) is non-negative, and (2) follows.

We now proceed with the proof of the theorem stated at the beginning. We may divide the interval $[a, b]$ into two parts corresponding to $f'(x) \leq 0$ and $f'(x) \geq 0$ and notice that it is clearly sufficient to prove the following theorem.

THEOREM. *If $f(x)$ is a real-valued, monotone, non-decreasing function possessing a second derivative $f''(x) \geq r > 0$ in $[a, b]$, then*

$$\left| \int_a^b \cos f(x) dx \right| \leq \int_0^{\beta_0} \cos(rx^2/2 + \mu_0) dx = \gamma_0/2r^{1/2}$$

where β_0, μ_0, γ_0 are the absolute constants defined above.

Proof. It is obviously no restriction to choose $a = 0$. Also, on placing $f(0) = \mu$, we may suppose, exactly as in the proof of the lemma, that

$$(7) \quad -\pi/2 \leq \mu < \pi/2.$$

Let

$$[0, b] = \sum_{j=0}^m [\delta_j, \delta_{j+1}]$$

be a subdivision of $[0, b]$ where the division points δ_j are determined by

$$(8) \quad \delta_0 = 0; \quad \delta_{m+1} = b; \quad f(\delta_j) = (j - 1/2)\pi \quad (j = 1, 2, \dots, m)$$

so that if x is in $[\delta_j, \delta_{j+1}]$ then

$$(j - 1/2)\pi \leq f(x) < (j + 1/2)\pi.$$

[If $f(b) \leq \pi/2$, we put $\delta_1 = b$, and the inequality (11) reduces to

$$\left| \int_0^b \cos f(x) dx \right| \leq \left| \int_0^{\delta_1} \cos f(x) dx \right|, \text{ which is obvious.}]$$

Since $f'(x)$ is non-negative and increasing throughout $[0, b]$, conditions (i)-(v) of the lemma are seen to be satisfied by the pair of functions

$$f_1(x) = f(x + \delta_{j+1} - \delta_j) - \pi; \quad f_2(x) = f(x)$$

in the pair of intervals

$$[a, b_1] = [\delta_j, \delta_{j+2} - \delta_{j+1} + \delta_j]; \quad [a, b_2] = [\delta_j, \delta_{j+1}],$$

for $(j = 1, 2, \dots, m - 2)$. Hence

$$(9) \quad \left| \int_{\delta_j}^{\delta_{j+1}} \cos f(x) dx \right| \geq \left| \int_{\delta_{j+1}}^{\delta_{j+2}} \cos f(x) dx \right| \quad (j = 1, 2, \dots, m - 2).$$

Although the lemma is not applicable in the case $j = m - 1$, it is quite obvious that (9) holds also for this value of j . In fact, we could extend the region of definition of $f(x)$ to $[0, \delta'_{m+1}]$ where

$$\delta'_{m+1} \geq \delta_{m+1} = b, \quad f(\delta'_{m+1}) = (m + 1/2)\pi;$$

and show

$$\left| \int_{\delta_m}^b \cos f(x) dx \right| \leq \left| \int_{\delta_m}^{\delta'_{m+1}} \cos f(x) dx \right| \leq \left| \int_{\delta_{m-1}}^{\delta_m} \cos f(x) dx \right|.$$

We now write

$$(10) \quad \int_0^b \cos f(x) dx = \int_0^{\delta_1} \cos f(x) dx + \sum_{j=1}^m \int_{\delta_j}^{\delta_{j+1}} \cos f(x) dx$$

and notice that the expression on the right is an alternating sum, since

$\cos f(x)$ is of different sign in two successive subintervals $[\delta_j, \delta_{j+1}]$ and $[\delta_{j+1}, \delta_{j+2}]$. Further, if we exclude the first term of this expression, the remaining m terms are of decreasing absolute value by (9). Hence, on placing

$$u_1 = \int_0^{\delta_1} \cos f(x) dx > 0, \quad u_2 = \int_{\delta_1}^{\delta_2} \cos f(x) dx < 0,$$

we obtain

$$u_1 \equiv \int_0^b \cos f(x) dx \geq u_1 + u_2 > u_2$$

and consequently

$$(11) \quad \left| \int_0^b \cos f(x) dx \right| \leq \max \left\{ \left| \int_0^{\delta_1} \cos f(x) dx \right|, \left| \int_{\delta_1}^{\delta_2} \cos f(x) dx \right| \right\}.$$

We now show that the conditions of the lemma are satisfied by the pair of functions

$$f_1(x) = f(x); \quad f_2(x) = rx^2/2 + \mu$$

in the pair of intervals

$$[a, b_1] = [0, \delta_1]; \quad [a, b_2] = [0, \beta], \quad \text{where } \beta = [(\pi - 2\mu)r]^{1/2},$$

if one chooses the integer k occurring in the lemma as zero. In fact, (i) is obviously satisfied since $f_1(0) = \mu$ by the definition of μ . Condition (ii) is satisfied since $f_1(\delta_1) = \pi/2$ from (8). The assumption of monotony of $f(x)$ insures conditions (iii), (iv) of the lemma, since $r > 0$. In order to verify that (v) also is satisfied we write

$$f_1''(x) = r + g(x),$$

so that $g(x) \geq 0$ in $[0, b]$ by assumption. This gives

$$f_1'(x) = rx + \int_0^x g(t) dt + f_1'(0),$$

where $f_1'(0) \geq 0$ by the monotony of $f(x)$ in $[0, b]$. Also

$$f_1(x) = rx^2/2 + \int_0^x \int_0^y g(t) dt dy + f_1'(0)x + \mu.$$

Thus the definition $f_2(x_2) = f_1(x_1)$ of x_2 leads to

$$x_2 = \left\{ x_1^2 + \frac{2}{r} \int_0^{x_1} \int_0^y g(t) dt dy + \frac{2x_1}{r} f_1'(0) \right\}^{1/2},$$

and by substituting this into $f_2'(x) = rx$ we check that

$$f_1(x_1) \geq f_2(x_2),$$

which is (v). The lemma then gives

$$(12) \quad \left| \int_0^{\delta_1} \cos f(x) dx \right| \leq \int_0^{\beta} \cos (rx/2 + \mu) dx,$$

where

$$\beta = [(\pi - 2\mu)/r]^{1/2}.$$

It is similarly shown that

$$(13) \quad \left| \int_0^{\delta_2 - \delta_1} \cos \{f(x + \delta_1) - \pi\} dx \right| \leq \int_0^{(2\pi/r)^{1/2}} \cos \left(\frac{rx^2 - \pi}{2} \right) dx.$$

The two inequalities (12) and (13) together with (11) give

$$(14) \quad \left| \int_0^b \cos f(x) dx \right| \leq \max H(\mu)/2r^{1/2}, \quad -\pi/2 \leq \mu < \pi/2$$

where $H(\mu)$ is independent of r , since

$$H(\mu) = 2\sqrt{r} \int_0^{\beta} \cos (rx^2/2 + \mu) dx, \quad \beta = [(\pi - 2\mu)/r]^{1/2}.$$

It may be shown by direct calculation of $H'(\mu)$ that there exists one and only one maximum point for $H(\mu)$ on the range $[-\pi/2, \pi/2]$, say at the point $\mu = \mu_0$ where μ_0 satisfies the equation

$$\int_0^{(\pi/2 - \mu_0)^{1/2}} \sin (x^2 + \mu_0) dx = 0.$$

Then if we choose $\mu = \mu_0$, the inequality (14) must hold for all functions $f(x)$ satisfying the conditions of the theorem, and

$$H(\mu_0) = 2\sqrt{2} \int_0^{(\pi/2 - \mu_0)^{1/2}} \cos (x^2 + \mu_0) dx$$

is the desired least permissible value γ_0 of all numbers γ satisfying (1).

SPECTRAL THEORY FOR A CERTAIN CLASS OF NON-SYMMETRIC COMPLETELY CONTINUOUS MATRICES.

By ANNA PELL WHEELER.

In this paper we consider the problem of the existence of characteristic numbers of a matrix A , which has real elements a_{ik} satisfying the condition

$$(\alpha) \quad a_{ik} = a_{ki}, \quad i, k > 1; \quad a_{i1} = -a_{i1}, \quad i > 1;$$

and which is completely continuous¹ with respect to the Hilbert space, i. e., the space of vectors $X = (x_1, x_2, \dots)$ such that $\sum |x_i|^2 < \infty$. A vector of the Hilbert space shall be called an *H-vector*. As is to be expected, the characteristic numbers are not all necessarily real nor simple poles of the resolvent. The convergence of expansions in terms of characteristic vectors is established for certain iterated matrices of A .

1. *Character and existence of characteristic numbers.* We shall denote by AX the linear transformation $(\sum_k a_{ik}x_k)$ of the vector $X = (x_1, x_2, \dots)$. So the system of linear equations

$$(1_1) \quad \lambda x_1 = a_{11}x_1 - \sum_k a_{k1}x_k,$$

$$(1_2) \quad \lambda x_i = a_{i1}x_1 + \sum_{k=2}^{\infty} a_{ik}x_k, \quad i > 1,$$

can be expressed by

$$(1) \quad \lambda X = AX.$$

If the equations

$$\begin{aligned} (2) \quad & \lambda X^{(1)} = AX^{(1)}, \\ & X^{(1)} + \lambda X^{(2)} = AX^{(2)}, \\ & \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ & \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ & X^{(r)} + \lambda X^{(r+1)} = AX^{(r+1)}, \end{aligned}$$

have solutions $X^{(1)}, X^{(2)}, \dots, X^{(r)}$ in the Hilbert space, but the $(r+1)$ -th equation does not have a solution in the Hilbert space, the characteristic number λ shall be said to be of *rank* r with respect to X_1 , or simply of *rank* r .

If the system (1) has a solution X for λ , the adjoint system

¹ Hellinger-Toeplitz, "Integralgleichungen und Gleichungen mit unendlichen Unbekannten," *Enzyklopädie der Mathematischen Wissenschaften*, Bd. II, Heft 9, p. 1400.

$$(3) \quad \lambda X^* = A'X^*$$

has a solution $X^* = (-x_1, x_2, \dots)$. The case in which a characteristic number of A is of rank > 1 with respect to X can arise only when

$$(4) \quad (X, X^*) = -x_1^2 + \sum_{k=2}^{\infty} x_k^2 = 0.$$

There can be only one characteristic number of rank > 1 . Assume that both λ and μ are of rank > 1 with respect to X and Y respectively. These vectors satisfy the conditions

$$(5) \quad (X, Y^*) = 0, \quad (X, X^*) = 0, \quad (Y, Y^*) = 0,$$

and from (5) it follows that

$$(6) \quad -(kx_1 + ly_1)^2 + \sum_{i=2}^{\infty} (kx_i + ly_i)^2 = 0$$

for all k and l . If now k and l are so chosen that the first term of (6) vanishes, then all the other terms must vanish, and the vectors X and Y would not be linearly independent, and hence could not belong to different characteristic numbers. A characteristic number cannot be of rank > 1 with respect to two different vectors X and Y , for again the conditions (5) would hold.

A characteristic number cannot be of rank > 3 . If the system (2) had solutions $X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}$, then $(X^{(1)}, X^{(1)*}) = 0$, $(X^{(2)}, X^{(1)*}) = 0$, and $(X^{(3)}, X^{(1)*}) = 0$. From the third equation of (2),

$$\begin{aligned} (X^{(2)*}, X^{(2)}) + \lambda(X^{(2)*}, X^{(3)}) &= (X^{(2)*}, AX^{(3)}) = (X^{(3)}, A'X^{(2)*}) \\ &= (X^{(3)}, X^{(1)*}) + \lambda(X^{(3)}, X^{(2)*}), \end{aligned}$$

and therefore $(X^{(2)}, X^{(2)*}) = 0$. It follows as in the two previous cases that $X^{(1)}$ and $X^{(2)}$ cannot be linearly independent. That a characteristic number may be of rank three is easily shown by special cases. There can be at most one pair of conjugate imaginary characteristic numbers. For, let $\lambda, \bar{\lambda}$ and $\mu, \bar{\mu}$ be two pairs of conjugate imaginary characteristic numbers, and let X, \bar{X} and Y, \bar{Y} be the corresponding characteristic vectors. Then $(X, \bar{X}^*) = 0$, $(Y, \bar{Y}^*) = 0$, $(X, \bar{Y}^*) = 0$, and if the constants $k = k_1 + ik_2$, $l = l_1 + il_2$ are chosen so that $kx_1 + ly_1 = 0$, the two vectors X and Y cannot be linearly independent. We have the following theorem.

THEOREM 1. *If the matrix A satisfies the condition (α) and is a completely continuous matrix with respect to the Hilbert space, there can be no more than one characteristic number of rank greater than unity, no characteristic number of rank greater than three, no characteristic number of rank greater than one with respect to more than one vector, and no more than one pair of conjugate imaginary characteristic numbers.*

For $i, k > 1$ the matrix (a_{ik}) is symmetric and completely continuous with respect to the Hilbert space and hence has characteristic² numbers μ_α , which are all real, and $\sum \mu_\alpha^2 < \infty$. Let L_α be the corresponding characteristic vectors, and we may assume that they form an orthonormal system, $(L_\alpha, L_\beta) = \delta_{\alpha\beta}$. The solution of (1₂), regarded as a non-homogeneous system in $x_i, i > 1$, is given by

$$(7) \quad \lambda x_i = x_1 a_{i1} + x_1 \sum_{\alpha} \frac{\mu_{\alpha} l_{\alpha i} \sum_2 l_{\alpha k} a_{k1}}{\lambda - \mu_{\alpha}}, \quad i > 1.$$

The substitution of (7) in (1) gives for $x_1 \neq 0$,

$$(8) \quad D(\lambda) = \lambda(\lambda - a_{11}) + \sum_2 a_{1i}^2 + \sum_{\alpha} \frac{\mu_{\alpha} \left(\sum_2 l_{\alpha i} a_{1i} \right)^2}{\lambda - \mu_{\alpha}} = 0,$$

a necessary and sufficient condition which λ must satisfy to be a characteristic number of A . The function $D(\lambda)$ is continuous in the intervals

$$\mu_{\alpha-1} + \epsilon \leq \lambda \leq \mu_{\alpha} - \epsilon, \quad \epsilon > 0$$

and changes sign between $\mu_{\alpha-1}$ and μ_{α} . Therefore there exists at least one characteristic number for A between two consecutive positive or two consecutive negative numbers μ_{α} . Hence the matrix A has real characteristic numbers except perhaps when the matrix $(a_{ik}), i, k > 1$, has only one or no characteristic numbers, or only one positive and one negative. In these three exceptional cases the equation (8) reduces to an algebraic equation. It can easily be shown that if the only root of (8) is zero, it is of rank greater than unity. In the sequel zero will be referred to as a characteristic number only if it is of rank greater than unity. Hence

THEOREM 2. *If a matrix A is not the zero matrix, satisfies the conditions (α) and is completely continuous with respect to the Hilbert space, then it has at least one characteristic number.*

2. Form of the expansion. If the vectors T_p and T_q belong to different characteristic numbers,

$$(9) \quad (T_p, T_q^*) = 0.$$

Whenever two vectors satisfy (9), and $(T_p, T_p^*) = 0$ or a negative value, then (T_q, T_q^*) must be positive, for otherwise constants c_1 and c_2 not both zero could be chosen so that $\sum_i (c_1 t_{pi} + c_2 t_{qi})^2$ would be zero or negative. Hence if T_p and T_q belong to a characteristic number of rank unity, it may be assumed that

² Hellinger-Toeplitz, *loc. cit.*, p. 1553.

$$(10) \quad (T_p, T_q^*) = \pm \delta_{pq},$$

where the minus sign holds for only one vector. Also if λ is of rank > 1 , linear combinations T_p of the solutions of (2) can be formed so that they satisfy (10).

If λ_a is a characteristic number of rank unity, let

$$(11) \quad E_p(X, X') = \lambda_p \sum_{a=1}^s \pm (T_p^{(a)}, X) (T_p^{(a)*}, X'),$$

where the $T_p^{(a)}$ form a fundamental system of characteristic vectors for λ_a , and the minus sign in (11) occurs before a term only if in that term the T_p satisfies $(T_p, T_p^*) = -1$.

If λ_i and $\bar{\lambda}_i$ form a pair of conjugate imaginary characteristic numbers, let

$$(12) \quad E_i(X, X') = \pm [\lambda_i (T^{(i)}, X) (T^{(i)*}, X') + \bar{\lambda}_i (\bar{T}^{(i)}, X) (\bar{T}^{(i)*}, X')],$$

where $(T^{(i)}, \bar{T}^{(i)*}) = 0$ and $(T^{(i)}, T^{(i)*}) = \pm 1$.

If λ_m is a characteristic number of rank $r > 1$, let

$$(13) \quad E_m(X, X') = \sum_{p,q=1}^r \pm c_{pq} (T_p^{(m)}, X) (T_q^{(m)*}, X'),$$

where $T_p^{(m)}$ are linear combinations of the solutions of (2) which satisfy (10); $AT_p^{(m)} = \sum_q c_{pq} T_q^{(m)}$; $c_{pq} = c_{qp}$; and the determinant $|c_{pq}| \neq 0$.

If $\lambda = 0$ is a characteristic number of rank $r > 1$, let

$$(14) \quad E_0(X, X') = \sum_{p,q=1}^r \pm c_{pq} (T_p^{(0)}, X) (T_q^{(0)*}, X'),$$

where the $T_p^{(0)}$ satisfy (10); $AT_p^{(0)} = \sum_q c_{pq} T_q^{(0)}$; $c_{pq} = c_{qp}$; and the determinant $|c_{pq}| = 0$.

The matrices E_p , E_i , E_m , E_0 are mutually orthogonal. In considering the expansion of $A(X, X')$ the question of convergence enters only if there is an infinite number of characteristic numbers of rank $= 1$, for then there would be an infinite series of the form

$$(15) \quad \sum_p E_p(X, X').$$

We shall now prove the following theorem.

THEOREM 3. *If the matrix A satisfies condition (α), is completely continuous with respect to the Hilbert space, and if the series (15) converges to a completely continuous bilinear form, then*

$$(16) \quad A(X, X') = \sum_p E_p(X, X') + E_i(X, X') + E_m(X, X') + E_0(X, X')$$

for any H -vectors X and X' .

Let $B(X, X')$ be the bilinear form formed by the difference of the two sides of (16). By hypothesis B is completely continuous, and the elements b_{ik} satisfy condition (α) . Hence by Theorem 2 if B is not the zero matrix, there exists at least one characteristic number. Let F_p, F_i, F_m, F_0, U_p bear the same relation to B as E_p, E_i, E_m, E_0, T_p do to A . The vectors T_p are orthogonal to B and therefore to all the vectors U_p , and hence to the F 's. On the other hand the U_p satisfy $AU_p = BU_p$ so that the U_p are linearly dependent on the T_p , and there is a contradiction. It follows that B is the zero matrix and the theorem is proved.

3. *Convergence of the expansion.* The iterated matrix $A^{(2)} = (\sum_j a_{ij} a_{jk})$ is completely continuous with respect to the Hilbert space, and satisfies the condition (α) . If $\lambda \neq 0$ is a characteristic number of rank r for A , then λ^2 is a characteristic number of rank r for $A^{(2)}$, and if λ is a characteristic number for $A^{(2)}$, with U_p as a characteristic vector, then $+\lambda^{1/2}$ or $-\lambda^{1/2}$, or both are characteristic numbers of A , with u or $u^{1/2} + Au$ as characteristic vectors. If $\lambda = 0$ is a characteristic vector of A of rank r , then $\lambda = 0$ is a characteristic number of rank $r - 1$ of $A^{(2)}$. The matrix $A^{(2)}$ is not the zero matrix unless $\lambda = 0$ is the only characteristic number of A and is of rank ≤ 2 . Similar results hold for the iterated matrices $A^{(n)}$; in particular zero is not a characteristic number of $A^{(n)}$ if $n > 3$.

THEOREM 4. *If the matrix A satisfies the condition (α) , and is completely continuous with respect to the Hilbert space, and if the vector*

$$(17) \quad X^{(0)} = (1, \sum_{i=2}^{\infty} \sum_p (a_{1i} l_{pi} l_{pk} / \mu_p))$$

does not satisfy

$$(18) \quad (X^{(0)}, X^{(0)*}) = 0,$$

and

$$(19) \quad AX^{(0)} = 0,$$

then the right hand side of

$$(20) \quad A^{(2)}(X, X') = \sum_p E_p^{(2)}(X, X') + E_i^{(2)}(X, X') + E_m^{(2)}(X, X') + E_0^{(2)}(X, X')$$

converges absolutely, and to the left hand side for any two H -vectors X and X' . If the vector $X^{(0)}$ does satisfy (18) and (19), then the right hand side of

$$(21) \quad A^{(4)}(X, X') = \sum_p E_p^{(4)}(X, X') + E_i^{(4)}(X, X') + E_m^{(4)}(X, X')$$

converges absolutely, and to the left hand side for any two H -vectors X and X' .

To prove the absolute convergence of (20), it is sufficient to show that the matrix³ $(\lambda_p t_{pi})$ is completely continuous. Only those terms in E_p for which

$$(22) \quad (T_p, T_p^*) = \sum_{i=2}^{\infty} t_{pi}^2 - t_{p1}^2 = 1$$

need be considered. And from (22) it follows that the matrix $(\lambda_p t_{pi})$ is completely continuous if t_{p1} is bounded, for then $\sum_i \lambda_p^2 t_{pi}^2$ converges. Assume first that the λ_p are all distinct from μ_α . The components of the vectors T_p are given by

$$(23) \quad t_{pi} = -\frac{t_{p1}}{\lambda_p} \left[a_{1i} + \sum_q \frac{\mu_q l_{qi} c_q}{\lambda_p - \mu_q} \right], \quad i > 1,$$

where $c_p = \sum_i l_{pi} a_{1i}$, and the substitution of (23) in (22) gives

$$(24) \quad t_{p1}^2 \left[-1 + \frac{\sum_2^{\infty} a_{1i}^2 - \sum_1^{\infty} c_q^2}{\lambda_p^2} + \sum_q \frac{c_q^2}{(\lambda_p - \mu_q)^2} \right] = 1.$$

The sequence t_{p1} is bounded unless

$$(25) \quad \sum_i a_{1i}^2 = \sum_p c_p^2,$$

and

$$(26) \quad \lim_{p' \rightarrow \infty} \sum_q \frac{c_q^2}{(\lambda_{p'} - \mu_q)^2} = 1,$$

for some subsequence $t_{p'1}$. We wish to show that under the conditions (25) and (26), the vector $X^{(0)}$ of (17) satisfies (18) and (19). The condition (25) says that the vector $X^{(0)}$ satisfies all except possibly the first equation of (19). From (26) we obtain

$$(27) \quad \sum_q \frac{c_q^2}{\mu_q^2} \leq 1,$$

and hence the absolute convergence of

$$(28) \quad \sum_q \frac{c_q^2}{\mu_q} \leq \sum_q c_q^2.$$

The equation (8) can now be written

$$(8') \quad D(\lambda_{p'}) = -a_{11} - \lambda_{p'} + \sum_q \frac{c_q^2}{\mu_q} + \lambda_{p'} \sum_q \frac{c_q^2}{(\lambda_{p'} - \mu_q) \mu_q} = 0.$$

³ For convenience the superscript α is omitted from $t_{pi}^{(\alpha)}$.

On account of (26) and (27) the coefficient of $\lambda_{p'}$ in the last term of (8') is bounded and therefore

$$(29) \quad a_{11} = \sum_q \frac{c_q^2}{\mu_q},$$

which expresses the fact that $X^{(0)}$ satisfies the first equation of (19). In virtue of (29) the equation (8') takes the form

$$(8'') \quad \lambda_{p'} \left(1 - \sum_q \frac{c_q^2}{(\lambda_{p'} - \mu_q)\mu_q} \right) = 0$$

and this combined with (26) and (27) gives

$$(30) \quad \sum_q \frac{c_q^2}{\mu_q^2} = 1,$$

which says that the vector $X^{(0)}$ satisfies (18). This completes the proof of the first part of the theorem.

For the second part of the theorem, it is sufficient to prove the complete continuity of the matrix $(\lambda_{p'}^2 t_{p'i})$ under the conditions (25) and (26). From (8'') and (30)

$$(31) \quad \sum_q \frac{c_q^2}{(\lambda_{p'} - \mu_q)\mu_q^2} = 0,$$

and hence

$$(32) \quad \sum_q \frac{c_q^2}{(\lambda_{p'} - \mu_q)^2} - 1 = \sum_q \frac{c_q^2}{(\lambda_{p'} - \mu_q)^2} - \sum_q \frac{c_q^2}{\mu_q^2} \\ = \lambda_{p'}^2 \sum_q \frac{c_q^2}{\mu_q^2 (\lambda_{p'} - \mu_q)^2} \geq \frac{\lambda_{p'}^2}{\sum_q c_q^2}$$

on account of (31) and (8'). And since $t_{p'1}^2$ is the reciprocal of (32) under (25) and (26), it follows that

$$\lambda_{p'}^2 t_{p'1}^2 \leq \sum_q c_q^2,$$

This shows that the matrix $(\lambda_{p'}^2 t_{p'i})$ is completely continuous, and the second part of the theorem is now a consequence of Theorem 3.

It remains to prove the theorem for the case in which an infinite number of λ 's are equal to μ_q . Let the other characteristic numbers of the matrix (a_{ik}) , $i, k > 1$, be denoted by μ_p . Since $c_q = \sum_2 l_{qi} a_{i1} = 0$, the system (1) has the solutions $V_p = (0, l_{p1}, l_{p2}, \dots)$ for which the matrix v_{qi} is bounded and $\lambda_q v_{qi}$ is completely continuous. If other solutions T_q exist for $\lambda = \mu_q$, the complete continuity of $\lambda_q t_{qi}$ and $\lambda_q^2 t_{qi}$ can be shown in the same way as above.

ON SUMMATION OF DERIVED SERIES OF THE CONJUGATE FOURIER SERIES.*

By A. F. MOURSUND.

1. *Introduction.* This paper is a continuation of an earlier paper in which we give theorems concerning the summability of the r -th, $r \geq 0$, derived series of the conjugate series of the Fourier series generated by a Lebesgue integrable function.¹ Our principal results are three theorems concerning the effectiveness of the N_{z_p} summation method² for the summation of the r -th derived series of the conjugate Fourier series. These theorems and the principal theorem of our earlier paper³ contain, as special cases, results, some of which are new, for the Bosanquet-Linfoot and Cesàro summation methods.⁴ The case $r = 0$ gives us three well known and one new theorem concerning the Cesàro summability of the conjugate Fourier series.

2. *Notation.* Throughout this paper we consider $f(x)$ to be Lebesgue integrable on $(-\pi, \pi)$ and periodic of period 2π . We use, as far as possible, the notation of I;⁵ and in addition set:

$$(2.1) \quad B_r(s) \equiv \int_0^s |A_r(s)|/s^r ds,$$

$$A^*_r(s) \equiv \int_0^s A_r(s) ds,$$

$$B^*_r(s) \equiv \int_0^s |A^*_r(s)|/s^{r+1} ds;$$

$$(2.2) \quad C^*_r \equiv \begin{cases} C_r & (r \text{ even}) \\ C_r - A^*_r(\pi) \frac{d^r}{ds^r} \cot s/2 \Big|_{s=\pi} & (r \text{ odd}); \end{cases}$$

* Presented to the American Mathematical Society, October 27, 1934.

¹ A. F. Moursund, "On summation of derived series of the conjugate Fourier series," *Annals of Mathematics* (2), vol. 36 (1935), pp. 182-193. Throughout this paper the paper cited here will be referred to as I.

² A. F. Moursund, "On the Nevanlinna and Bosanquet-Linfoot summation methods," *Annals of Mathematics* (2), vol. 35 (1934), pp. 239-247 (Section II).

³ *Loc. cit.*², Theorem 10.1.

⁴ The question of inclusion of other methods by the N_{z_p} method is discussed in *loc. cit.*², Section II.

⁵ See § 2. To avoid confusion, the functions there designated by $A^*_r(s)$ and $B_r(n)$ should be designated by other letters, and the functions $F_i(ns)$ and $G_i(ns)$ should be designated, respectively, as $F(i, r+1, ns)$ and $G(i, r+1, ns)$.

$$(2.3) \quad D_r^*(n) \equiv \begin{cases} D_r(n) & (r \text{ even}) \\ D_r(n) - A_r^*(\pi) \left[\sum_{j=0}^{(r-1)/2} \frac{r!}{(2j+1)!(r-2j-1)!} \frac{d^{2j+1}}{ds^{2j+1}} \cot s/2 \right. \\ \quad \times \left. \frac{d^{r-2j-1}}{ds^{r-2j-1}} \cos ns + \frac{dr}{ds^r} \sin ns \right]_{s=\pi} & (r \text{ odd}); \end{cases}$$

$$(2.4) \quad \tilde{f}^{*(r)}(x) \equiv -\lim_{\epsilon \rightarrow 0} \int_{\epsilon}^{\pi} A_r^*(s) \frac{d^{r+1}}{ds^{r+1}} \cot s/2 ds - C_r^*;$$

and

$$(2.5) \quad F(i, p, v) \equiv \int_0^1 z_p(t) t^i \frac{dt}{dt^i} \cos vt dt,$$

$$G(i, p, v) \equiv \int_0^1 z_p(t) t^i \frac{dt}{dt^i} \sin vt dt.$$

3. *Expressions for $\sigma_n^{(r)}(x)$ and $N_{z, \sigma_n^{(r)}}(x)$.* In § 3 of I we showed that

$$(3.1) \quad \sigma_n^{(r)}(x) = \int_0^{\pi} A_r(s) \frac{dr}{ds^r} \theta(s, n) ds - C_r + D_r(n).$$

Proceeding as we did there in deriving (3.6), we have

$$(3.2) \quad N_{z, \sigma_n^{(r)}}(x) = \int_{1/n^2}^{\pi} A_r(s) \frac{dr}{ds^r} \cot s/2 ds - C_r$$

$$+ \int_0^1 z_r(t) dt \int_0^{1/n^2} A_r(s) \frac{dr}{ds^r} \theta(s, nt) ds$$

$$- \sum_{i=0}^{r-1} \frac{r!}{i!(r-i)!} \left[\int_{1/n^2}^{\delta} + \int_{\delta}^{\pi} \right] A_r(s)/s^i \frac{d^{r-i}}{ds^{r-i}} \cot s/2 F(i, r, ns) ds$$

$$- \left[\int_{1/n^2}^{A/n} + \int_{A/n}^{\delta} + \int_{\delta}^{\pi} \right] A_r(s)/s^r \cot s/2 F(r, r, ns) ds$$

$$+ \left[\int_{1/n^2}^{A/n} + \int_{A/n}^{\delta} + \int_{\delta}^{\pi} \right] A_r(s)/s^r G(r, r, ns) ds$$

$$+ \int_0^1 z_r(t) D_r(nt) dt$$

$$\equiv \int_{1/n^2}^{\pi} A_r(s) \frac{dr}{ds^r} \cot s/2 ds - C_r + \sum_{i=1}^{10} J_i.$$

Upon integrating (3.1) by parts we obtain, with the aid of Lemmas 3.1 and 3.2 of I,

$$(3.3) \quad \sigma_n^{(r)}(x) = - \int_0^{\pi} A_r^*(s) \frac{d^{r+1}}{ds^{r+1}} \theta(s, n) ds - C_r^* + D_r^*(n).$$

Our use of (3.3) is analogous to our use of (3.1).

4. *Lemmas.* In proving our theorems we use, in addition to the lemmas of I, the following lemmas.⁶

LEMMA 4.1. For v on $(0, \infty)$ when $0 \leq i < p$ and on $(0, A)$ when $i = p$

$$F(i, p, v) = P_{i_1}(v) - P_{i_2}(v)$$

$$G(i, p, v) = Q_{i_1}(v) - Q_{i_2}(v)$$

where the P 's and Q 's are bounded monotone decreasing positive functions of v .

LEMMA 4.2. For A sufficiently large and $v \geq A$,

$$|F(p, p, v)|, |G(p, p, v)| \leq K/v + Z_p(v)$$

where K is a constant.

Proof. For v sufficiently large

$$\begin{aligned} |F(p, p, v)| &= \left| \int_0^1 z_p(t) t^p \frac{d^p}{dt^p} \cos vt \, dt \right| \\ &= \left| \int_0^1 \frac{d^p}{dt^p} [z_p(t) t^p] \cos vt \, dt \right| \leq K/v + Z_p(v). \end{aligned}$$

The proof is similar for $G(p, p, v)$.

LEMMA 4.3. $F(p, p, v)$ and $G(p, p, v) \rightarrow 0$ as $v \rightarrow \infty$.

The functions which we consider in the remaining lemmas of this section are defined by (3.2).

LEMMA 4.4. Wherever $\tilde{f}^{(r)}(x)$ exists J_n and $J_r \rightarrow 0$ as $n \rightarrow \infty$ for a fixed A .

LEMMA 4.5. At a point where $A_r(s) = 0(s^r)$, as $s \rightarrow 0$, J_5 and $J_8 \rightarrow 0$ as $A, n \rightarrow \infty$ for a fixed sufficiently small δ .

Proof. Using Lemma 4.2 we have, for a fixed sufficiently small δ ,

⁶ In proving the lemmas we use properties of the N_{z_p} method, *loc. cit.* 2, Section II, without further referring to them. The proof of Lemma 4.1 is like the proof of Lemma 6.3 and (6.1) of I; that of Lemma 4.4 like that of Lemma 8.2 of I. In the proof of Lemma 4.5 the last term $\rightarrow 0$ as $A, n \rightarrow \infty$ because $\int_C^\infty Z_r(s)/s^r \, ds$ exists.

$$\begin{aligned}
|J_5| &\leq 2 \int_{A/n}^{\delta} |A_r(s)|/s^r [K/ns^2 + Z_r(ns)/s] ds \\
&= 0 \left[\frac{1}{n} \int_{A/n}^{\delta} 1/s^2 ds + \int_{A/n}^{\delta} Z_r(ns)/s ds \right] \\
&= 0 \left[1/A + \int_A^{n\delta} Z_r(s)/s ds \right] = o(1) \text{ as } A, n \rightarrow \infty.
\end{aligned}$$

Similarly $J_8 \rightarrow 0$.

LEMMA 4.6. *At a point where $B_r(s) = o(s)$, as $s \rightarrow 0$, J_5 and $J_8 \rightarrow 0$ as $\delta \rightarrow 0$ and $n \rightarrow \infty$ for a fixed sufficiently large A .*

Proof. Integrating by parts we have

$$\begin{aligned}
|J_5| &\leq 2 \int_{A/n}^{\delta} |A_r(s)|/s^r [K/ns^2 + Z_r(ns)/s] ds \\
&= 2B_r(s)/s [K/ns + Z_r(ns)] \Big|_{A/n}^{\delta} + 2K/n \int_{A/n}^{\delta} B_r(s)/s \cdot 1/s^2 ds \\
&\quad - 2 \left[\int_{A/n}^{\delta} B_r(s)/s Z'_r(ns) ds + \int_{A/n}^{\delta} B_r(s)/s Z_r(ns)/s ds \right] \\
&= o(1) + o(1) + o(1)Z_r(ns) \Big|_{A/n}^{\delta} + o(1) \int_{A/n}^{\delta} Z_r(ns)/s ds \\
&= o(1) \text{ as } \delta \rightarrow 0 \text{ and } n \rightarrow \infty. \text{ Similarly } J_8 = o(1).
\end{aligned}$$

LEMMA 4.7. J_6 and $J_9 \rightarrow 0$ as $n \rightarrow \infty$ for a fixed δ .

Proof. The lemma is a consequence of Lemma 4.3.

5. *Theorems concerning the existence of $\bar{f}^{*(r)}(x)$ and the behavior of the functions defined by (2.1).* In § 9 of I we gave sufficient conditions for the existence of $\bar{f}^{(r)}(x)$.⁷ We consider here some further consequences of these conditions.

THEOREM 5.1. *If $f(x)$ is such that $d^{r-1}f(x)/dx^{r-1}$ is of bounded variation on $(-\pi, \pi)$ then, as $s \rightarrow 0$, $A_r(s) = O(s^r)$ and $B_r(s) = o(s)$ almost everywhere.*

Proof. By Lemma 7.6 of I $A_r(s) = O(s^r)$ when $d^r f(x)/dx^r$ exists. The theorem follows for when $d^{r-1}f(x)/dx^{r-1}$ is of bounded variation $d^r f(x)/dx^r$ exists almost everywhere.

⁷ In Lemma 7.8 and Theorem 9.3 of I the integral $\int_{\epsilon}^{\pi} dr/dsr \Phi_r(s) \cot s/2 ds$ should be replaced by $\int_{\epsilon}^{\pi} \cot s/2 d[dr^{-1}/dsr^{-1} \Phi_r(s)]$; the existence almost everywhere of the limit as $\epsilon \rightarrow 0$ depends on Plessner's proof of the existence of $\bar{f}(x)$ rather than on his proof of the existence of $\bar{f}(x)$.

THEOREM 5.2. *Wherever the generalized derivative $f^{(r+1)}(x)$ exists $A_r(s) = O(s^r)$ and $B_r(s) = o(s)$, as $s \rightarrow 0$.*

Proof. Wherever $f^{(r+1)}(x)$ exists we see using Definition (2.1) of I that

$$A_r(s) = (-1)^{r+1}/\pi [f^{(r+1)}(x) + w(x, s)]s^{r+1}$$

where $w(x, s) \rightarrow 0$ as $s \rightarrow 0$. The theorem follows.⁸

THEOREM 5.3. *The existence of $\tilde{f}^{(r)}(x)$ implies the existence of $\tilde{f}^{*(r)}(x)$, but the converse is not necessarily true.*⁹

THEOREM 5.4. *$A_r^*(s) = O(s^{r+1})$ wherever $A_r(s) = O(s^r)$ and $B_r^*(s) = o(s)$ wherever $B_r(s) = o(s)$, but the converse is not necessarily true.*

Proof. Wherever $B_r(s) = o(s)$

$$B_r^*(s) \leq \int_0^s 1/s^{r+1} ds \int_0^s |A_r(t)| dt \leq \int_0^s B_r(s)/s ds = o(s).$$

The rest of the proof is left to the reader.

6. *Summability theorems.* In this section we give our principal theorems and, for convenience in reference, state also the principal theorem of I. We assume, as everywhere in this paper, that the function $f(x)$ which generates the series under consideration is Lebesgue integrable on $(-\pi, \pi)$ and of period 2π .

THEOREM 6.1. *The N_{z_p} sum of the r -th ($r = 0, 1, 2, \dots, p$), derived series of the conjugate Fourier series is $\tilde{f}^{(r)}(x)$ wherever that limit exists and, as $s \rightarrow 0$, either $A_r(s) = O(s^r)$ or $B_r(s) = o(s)$.*

THEOREM 6.2. *The N_{z_p} sum of the r -th ($r = 0, 1, 2, \dots, p-1$), derived series of the conjugate Fourier series is $\tilde{f}^{(r)}(x)$ wherever that limit exists.*

Proof of Theorems 6.1 and 6.2. Theorem 6.2 is Theorem 10.1 of I. Referring to the proof of that theorem and (3.2) we see that at a point where $\tilde{f}^{(r)}(x)$ exists J_1, J_2, J_3 , and $J_{10} \rightarrow 0$, respectively, like K_1, K_2, K_3 , and $K_6 \rightarrow 0$. When, also, either $A_r(s) = O(s^r)$ or $B_r(s) = o(s)$ it follows from

⁸ In fact $A_r(s) = O(s^{r+1})$ and $B_r(s) = O(s^2)$.

⁹ See B. N. Prasad, "Contribution à l'étude de la Série Conjuguée d'une Série de Fourier," *Journal de Mathématiques Pures et Appliquées* (9), vol. 11 (1932), pp. 153-205 (p. 178).

Lemmas 4. 4, 4. 5, 4. 6, and 4. 7 that J_4 to J_9 , inclusive, $\rightarrow 0$. Whence, as do corresponding functions in the proof of Theorem 10. 1,

$$N_{z_r \sigma_n^{(r)}}(x) \text{ and } N_{z_p \sigma_n^{(r)}}(x), \quad p > r, \rightarrow \bar{f}^{(r)}(x) \text{ as } n \rightarrow \infty.$$

This completes the proof of Theorem 6. 1.

THEOREM 6. 3. *The N_{z_p} sum of the r -th ($r = 0, 1, 2, \dots, p-1$), derived series of the conjugate Fourier series is $\bar{f}^{*(r)}(x)$ wherever that limit exists and, as $s \rightarrow 0$, either $A^*_r(s) = 0(s^{r+1})$ or $B^*_r(s) = o(s)$.*

THEOREM 6. 4. *The N_{z_p} sum of the r -th ($r = 0, 1, 2, \dots, p-2$), derived series of the conjugate Fourier series is $\bar{f}^{*(r)}(x)$ wherever that limit exists.*

Proof of Theorems 6. 3 and 6. 4 are similar to those of Theorems 6. 1 and 6. 2. We first form from (3. 3) expressions for the $N_{z_{r+1}}$ and $N_{z_{r+2}}$ transforms of $\sigma_n^{(r)}(x)$ which are similar, respectively, to (3. 2) of this paper and (3. 6) of I; next prove lemmas similar to the lemmas of § 5 of I, replacing $A_r(s)$ by $A^*_r(s)$, $\frac{d^r}{ds^r} \cot s/2$ by $\frac{d^{r+1}}{ds^{r+1}} \cot s/2$, et cetera; and then proceed as in the proofs of Theorems 6. 1 and 6. 2.¹⁰

7. *Theorems for Bosanquet-Linfoot and Cesàro summability.* By choosing the functions

$$(7. 1) \quad B_{\alpha, \beta}(t) = G(1-t)^{\alpha-1}(\log C/1-t)^{-\beta}$$

with $\alpha > p$ or $\alpha = p$, $\beta > 1$, and $G = G(\alpha, \beta, C)$ such that

$$\int_0^1 B_{\alpha, \beta}(t) dt = 1;$$

and

$$(7. 2) \quad (p + \delta)(1-t)^{p+\delta-1}$$

with $\delta > 0$, respectively, as kernels for the N_{z_p} method we obtain theorems for the Bosanquet-Linfoot and Cesàro methods from the theorems of § 6.¹¹

¹⁰ As a consequence of the Riemann-Lebesgue theorem Lemma 6. 5 of I remains valid when $z_{r+1}(t)$ is replaced by $z_r(t)$. Hence terms arising from $D^*_{r+1}(n) \rightarrow 0$ like those arising from $D_r(n)$.

¹¹ See *loc. cit.* ², Section II. The Cesàro method theorems are obtained as a result of the equivalence of the Riesz and Cesàro methods.

Among the theorems thus obtained are three well known theorems concerning the Cesàro summability of the conjugate Fourier series¹² and the following new theorem.

THEOREM 7.1. *The conjugate Fourier series is summable $(C, 2 + \delta)$, for every $\delta > 0$, to*

$$\bar{f}^*(x) = \lim_{\epsilon \rightarrow 0} \frac{1}{4\pi} \int_{\epsilon}^{\pi} \left\{ \int_0^s [f(x+t) - f(x-t)] dt \right\} \csc^2 s/2 ds$$

wherever that limit exists.

Theorem 7.1 is related to the third of the Cesàro method theorems mentioned above (Prasad's theorem) like the second is to the first.

THE UNIVERSITY OF OREGON,
EUGENE, OREGON.

¹² See E. W. Hobson, *The Theory of Functions of a Real Variable*, vol. II (2nd Edition), p. 697, for the first of these theorems which is the standard (C, δ) , $\delta > 0$, summability theorem for the conjugate Fourier series; for the second see R. E. A. C. Paley, "On the Cesàro summability of Fourier series and allied series," *Proceedings of the Cambridge Philosophical Society*, vol. 26 (1930), pp. 173-203 (Theorem 2 with $\alpha = 0$); and for the third see Prasad, *loc. cit.* ⁹, Theorem IV.

A CERTAIN MIXED LINEAR INTEGRAL EQUATION.

By OLIVE MARGARET HUGHES.

Introduction. The purpose of this paper is the development of a spectral theory for the linear functional transformation

$$(1) \quad Tf(x) = MK(x, \xi)f(\xi) + \int_0^1 K(x, s)f(s)ds$$

of real-valued functions $f(x)$, continuous on the interval $I(0 \leq x \leq 1)$, where

(a) The kernel $K(x, s)$ is real, symmetric¹ and continuous on the square $R(0 \leq x \leq 1, 0 \leq s \leq 1)$.

(b) ξ is a given fixed value of the variable of integration in I .

(c) M is a known negative constant.

W. A. Hurwitz² has obtained some alternative theorems for such transformations. For M positive or zero the above problem reduces respectively to a Kneser weighted problem³ and to the well-known Fredholm problem; more recent works⁴ include both of these as special cases of a general spectral theory for linear symmetric functional transformations. They do not cover the case however with which we are concerned, since under the given hypothesis characteristic numbers of the integral equation

$$(2) \quad \lambda u(x) = MK(x, \xi)u(\xi) + \int K(x, s)u(s)ds$$

may be imaginary or zero, and may be poles of the resolvent of order greater than one. The following simple examples illustrate these possibilities:

¹ A definition of symmetric functions is given on page 871.

² W. A. Hurwitz, "Mixed linear integral equations of the first order," *Transactions of the American Mathematical Society*, vol. 16 (1915).

³ A. Kneser, *Die Integralgleichungen*, Rendiconti, vol. 14 (1914). Also his *Die Integralgleichungen und ihre Anwendungen in der Mathematischen Physik*, 2 umgearbeitete Auflage (1922).

⁴ J. von Neumann, "Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren," *Mathematische Annalen*, vol. 102. Also M. H. Stone, "Linear transformations in Hilbert space and their applications to analysis," *Colloquium Lectures of the American Mathematical Society*, vol. 15 (1932).

Ex. 1. If $K(x, s) = xs$, $M = -1/3$, $\xi = 1$, then $u(x) = x$ is a characteristic function corresponding to the characteristic number $\lambda = 0$ of the first order.

Ex. 2. If $K(x, s) = 1 - 4xs$, $M = -1/3$, $\xi = 1$, then $u(x) = 1 - 2x$ is a characteristic function corresponding to $\lambda = 1/3$ of order two.

Ex. 3. If $K(x, s) = 1 - xs$, $M = -1$, $\xi = 1$, then $u(x) = [(-5)^{1/2} - 2] + 3x$ and $\bar{u}(x) = -[(-5)^{1/2} + 2] + 3x$ are conjugate imaginary characteristic functions corresponding respectively to the conjugate imaginary characteristic numbers $\lambda = [(-5)^{1/2} + 2]/6$, $\bar{\lambda} = -[(-5)^{1/2} - 2]/6$.

We adopt the notation

$$\int^* g(s) ds = Mg(\xi) + \int g(s) ds,$$

thus the transformation (1) and integral equation (2) are now written

$$Tf(x) = \int^* K(x, s)f(s) ds$$

$$\lambda u(x) = \int^* K(x, s)u(s) ds.$$

The operator \int^* we name the *M-integral*, and for convenience we name the corresponding operation *M-integration*. The *M-integral* is a special instance of the Stieltjes integral, and we shall assume its elementary properties throughout this paper without enumerating them.

In Section I we begin by defining an *M-orthonormed* system of functions, that is, orthonormed with respect to the *M-integral*, and which reduces, for $M = 0$, to an ordinary orthonormed system. In laying down this definition allowance must be made for the fact that the *M-integral* of a positive function may be negative, and it is on account of this property that we are unable to generalize the fundamental inequalities of Bessel and Schwarz. In the same section we establish the existence of a complete *M-orthonormed* system.

The complete *M-orthonormed* system forms the starting point for the solution of our problem. We use it to pass from the integral equation (2), following the classic method of Hilbert, to a system of equations in infinitely many variables, which has been solved by A. Pell-Wheeler.⁵ This furnishes a proof of the existence of characteristic numbers for the integral equation, and also an expansion theorem. The results are given in Section II.

⁵ A. Pell-Wheeler, "Spectral theory for a certain class of non-symmetric completely continuous matrices," *American Journal of Mathematics*, vol. 57 (1935), pp. 847-853.

SECTION I. *M*-ORTHONORMED SYSTEMS.

1. *Preliminary definitions.* The functions

$$\phi_1(x), \phi_2(x), \dots, \phi_n(x), \dots,$$

finite or infinite in number, form an *M-orthogonal system* if they satisfy the conditions

$$\int^* \phi_i(x) \phi_j(x) dx = 0, \quad i \neq j.$$

They form an *M-orthonormed system* if

$$\int^* \phi_i(x) \phi_j(x) dx = \pm e_{ij}, \quad e_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j. \end{cases}$$

A system of functions $\{\phi_n(x)\}$, continuous on I , is *M-closed* if there exists no function $f(x)$, continuous on I , which is *M-orthogonal* to all the functions of the system.

An *M-orthonormed system* of functions $\{\phi_n(x)\}$, continuous on I , is *M-complete* if, for all possible functions $f(x)$ continuous on I , the functions ϕ_n of the system satisfy the relation

$$\sum_{n=1}^{\infty} \pm \left(\int^* f \phi_n \right)^2 = \int^* f^2,$$

the negative sign in the summation occurring when and only when the value of n is such that $\int^* \phi_n^2 = -1$.

2. *Construction of an M-orthonormed system.* Let $\{f_n(x)\}$ be a set of functions, infinite in number, which satisfy the conditions

- (a) They are continuous on I .
- (b) They are linearly independent on I .
- (c) At least one function of the set does not have the value zero at $x = \xi$.
- (d) $D_n(f) \neq 0, \quad (n = 1, 2, \dots),$

where $D_n(f)$ is the n -rowed determinant with general element $\int^* f_i f_j$ ($i, j = 1, 2, \dots, n$).

From this set of functions we build up by the method of determinants an *M-orthonormed system* $\{\phi_n(x)\}$ as follows:

$$(1a) \quad \begin{cases} \phi_1 = f_1 / \sqrt{\pm \int^* f_1^2} \\ \phi_n = E_n(f) / \sqrt{\pm D_{n-1}(f) \cdot D_n(f)}, \quad (n = 2, 3, \dots), \end{cases}$$

where $E_n(f)$ is the determinant derived from $D_n(f)$ by replacing the elements of the n -th column, reading downwards, by f_1, f_2, \dots, f_n . The positive (negative) sign is to be used if the product of the determinants under the radical sign is positive (negative), and in the case of $n = 1$ if $\int^* f_1^2$ is positive (negative).

It can be easily verified that the system $\{\phi_n\}$ thus defined satisfies the definition of an M -orthonormed system. Each ϕ_n is expressed linearly and homogeneously, with constant coefficients, in terms of the functions f_1, f_2, \dots, f_n , and conversely.

The functions ϕ_n are real and continuous on I . They are furthermore linearly independent there; for, suppose there exist constants c_n , not all zero, such that

$$\sum_{n=1}^p c_n \phi_n \equiv 0;$$

multiplication throughout by ϕ_n and M -integration over I leads to $c_n = 0$ ($n = 1, 2, \dots, p$), a contradiction.

Conversely, from an M -orthonormed system of continuous functions $\{\phi_n\}$ it is possible to construct an orthonormed set $\{f_n\}$ of the form

$$f_n = \sum_{p=1}^n c_p \phi_p$$

where the c_p are constants, not all zero. This property is an immediate consequence of the linear independence of the ϕ_n , just proved.

The f_n may be transformed into an orthonormed system of functions which satisfy the four conditions specified above, and since such a transformation simplifies considerably the expressions for ϕ_n we shall assume it has been previously carried through; then formula (1a) reduces, after expansion of the determinants to the alternative form

$$(1b) \quad \begin{cases} \phi_1 = \frac{f_1}{\sqrt{\pm (1 + M f_{1,\xi}^2)}} \\ \phi_n = \frac{f_n (1 + M \sum_{i=1}^{n-1} f_{i,\xi}^2) - M f_{n,\xi} \sum_{i=1}^{n-1} f_{i,\xi} f_i}{\sqrt{\pm (1 + M \sum_{i=1}^{n-1} f_{i,\xi}^2) (1 + M \sum_{i=1}^n f_{i,\xi}^2)}}, \quad (n = 2, 3, \dots). \end{cases}$$

The subscript ξ indicates that the value of the function is to be taken at $x = \xi$.

It is desirable at this point to introduce the following abbreviations:

$$A_p(x) = \sum_{i=1}^p f_{i,\xi} f_i(x), \quad A_0(x) \equiv 0,$$

$$P_p = (1 + MA_{p-1,\xi})(1 + MA_{p,\xi}), \quad (p = 1, 2, \dots).$$

Then formula (1a) assumes the final abridged form

$$(1c) \quad \phi_n = \frac{f_n(1 + MA_{n-1,\xi}) - Mf_{n,\xi}A_{n-1}}{\sqrt{\pm P_n}}, \quad (n = 1, 2, \dots).$$

We are able to make certain general deductions concerning the sign of P_n . There are two possibilities:

(a) P_n is positive for all n . Then $1 + MA_{n,\xi}$ is also, since otherwise P_1 would be negative. Whence

$$A_{n,\xi} < -1/M, \quad (n = 1, 2, \dots).$$

Since $A_{n,\xi}$ is monotonic, increasing and bounded, $\lim_{n \rightarrow \infty} A_{n,\xi}$ exists and is finite. This is a necessary condition that $P_n > 0$ ($n = 1, 2, \dots$).

(b) P_n is positive for all values of n except one, say N , which satisfies the inequality

$$-1/A_{N-1,\xi} < M < -1/A_{N,\xi}, \quad N > 1$$

or

$$M < -1/A_{N,\xi}, \quad N = 1.$$

In this case P_N is negative. A sufficient condition for the existence of such a value N of n is that $\lim_{n \rightarrow \infty} A_{n,\xi} = \infty$.

P_n cannot be negative for more than one value of n under any circumstances, for if N is the first value of n for which $P_n < 0$, we have

$$1 + MA_{N,\xi} < 0,$$

$$\therefore 1 + MA_{n,\xi} < 0, \quad n > N,$$

$$\therefore P_n = (1 + MA_{n-1,\xi})(1 + MA_{n,\xi}) > 0, \quad n > N.$$

3. *M-completeness of the system (1c).* From the foregoing results we obtain

THEOREM I. *If the functions f_n satisfy the four conditions of § 2, then the set $\{\phi_n\}$ given by (1c) can contain at most one function satisfying the relation*

$$\int^* \phi_i^2 = -1.$$

In order to facilitate the proof of the next theorem we first introduce the

LEMMA. *If the set $\{f_n\}$ is complete, then a necessary and sufficient condition that*

$$(2) \quad \lim_{n \rightarrow \infty} \frac{\left(\int f A_n - f_\xi \right)^2}{1 + M A_{n,\xi}} = 0$$

is that $\lim_{n \rightarrow \infty} A_{n,\xi} = \infty$, where f is an arbitrary function continuous on I .

(a) *Necessary condition.* From the Lagrange-Cauchy inequality we find

$$\left| \int f A_n \right| \leq \sum_{i=1}^n \left| f_{i,\xi} \int f f_i \right| \leq \sqrt{A_{n,\xi}} \cdot \sqrt{\sum_{i=1}^n \left(\int f f_i \right)^2}$$

for all n , and all continuous functions f , and therefore

$$\sqrt{A_{n,\xi}} \geq \left| \int f A_n \right| / \sqrt{\sum_{i=1}^n \left(\int f f_i \right)^2}.$$

If $\lim_{n \rightarrow \infty} A_{n,\xi}$ is finite, then from (2), $\lim_{n \rightarrow \infty} \int f A_n = f_\xi$ for all f under consideration, and hence

$$\lim_{n \rightarrow \infty} \left| \int f A_n \right| / \sqrt{\sum_{i=1}^n \left(\int f f_i \right)^2} = |f_\xi| / \sqrt{\int f^2}.$$

For a suitable choice of f this limit can be made greater than any arbitrarily assigned positive number,⁶ which contradicts the assumption that $\lim_{n \rightarrow \infty} A_{n,\xi}$ be finite.

⁶ For example, define $f(x) = F_n^{-2}(x)$ as follows:

$$\begin{aligned} F_n^{-2}(x) &= 0, & 0 \leq x \leq (n-1)\xi/n \\ &= nx/\xi + (1-n), & (n-1)\xi/n \leq x \leq \xi \\ &= nx/(\xi-1) + (\xi-n\xi-1)/(\xi-1), & \xi \leq x \leq (n\xi+1-\xi)/n \\ &= 0, & (n\xi+1-\xi)/n \leq x \leq 1, \end{aligned}$$

$\xi \neq 0$ or 1 . As n increases, $\int f^2$ decreases, but $f(\xi) = 1$ for all n . If $\xi = 0$ or 1 this definition needs to be modified somewhat.

(b) *Sufficient condition.* If $\int fA_n$ approaches a finite limit with increasing n , the sufficiency of the condition is self-evident; it remains then, to investigate the case where this limit is infinite. To this end it suffices to prove that

$$\lim_{n \rightarrow \infty} \int fA_n / \sqrt{A_{n,\xi}} = 0.$$

This fraction may be written in the form

$$\int fA_n / \sqrt{A_{n,\xi}} = \sum_{i=1}^m f_{i,\xi} \int ff_i / \sqrt{A_{n,\xi}} + \sum_{i=m+1}^n f_{i,\xi} \int ff_i / \sqrt{A_{n,\xi}}, \quad n > m.$$

Application of the Lagrange-Cauchy inequality to the second term on the right gives

$$\left| \sum_{i=m+1}^n f_{i,\xi} \int ff_i / \sqrt{A_{n,\xi}} \right| \leq \left[\sqrt{\sum_{i=m+1}^n f_{i,\xi}^2} / \sqrt{A_{n,\xi}} \right] \cdot \sqrt{\sum_{i=m+1}^n \left(\int ff_i \right)^2} \\ \leq 1 \cdot \epsilon/2 \text{ for } m > N_1, n > m.$$

If we take a fixed value for $m > N_1$, then

$$\left| \sum_{i=1}^m f_{i,\xi} \int ff_i / \sqrt{A_{n,\xi}} \right| \leq \left[\max |f_{i,\xi}| \cdot \max \left| \int ff_i \right| / \sqrt{A_{n,\xi}} \right] \cdot m, \\ (i = 1, 2, \dots, m). \\ \leq \epsilon/2 \text{ for } n > N_2 > N_1.$$

It follows that

$$\left| \int fA_n / \sqrt{A_{n,\xi}} \right| = \left| \sum_{i=1}^n f_{i,\xi} \int ff_i / \sqrt{A_{n,\xi}} \right| < \epsilon, \quad n \geq N_2,$$

which completes the proof of the Lemma.

THEOREM II. *If the set $\{f_n\}$ of § 2 is complete, then a necessary and sufficient condition that the system $\{\phi_n\}$ defined by (1c) be M-complete is that*

$$\lim_{n \rightarrow \infty} A_{n,\xi} = \infty.$$

If we substitute the given expression for ϕ_n in the series

$$\sum_{n=1}^{\infty} \pm \left(\int^* f \phi_n \right)^2$$

we get

$$\begin{aligned} \sum_{n=1}^{\infty} \pm \left(\int^* f \phi_n \right)^2 &= \sum_{n=1}^{\infty} \left[(1 + MA_{n-1, \xi}) \int^* ff_n - Mf_{n, \xi} \int^* fA_{n-1} \right]^2 / P_n \\ &= \sum_{n=1}^{\infty} \left[M^2 f_{n, \xi}^2 f_{\xi}^2 + 2Mf_{n, \xi} f_{\xi} \left\{ (1 + MA_{n-1, \xi}) \int ff_n - Mf_{n, \xi} \int fA_{n-1} \right\} \right. \\ &\quad \left. + \left\{ (1 + MA_{n-1, \xi}) \int ff_n - Mf_{n, \xi} \int fA_n \right\}^2 / P_n \right]. \end{aligned}$$

The sum S_p of the first p terms of this series is

$$\begin{aligned} S_p &= \left[M^2 A_{p, \xi} f_{\xi}^2 + 2Mf_{\xi} \int fA_p + \sum_{i=1}^p \left(\int ff_i \right)^2 (1 + MA_{p, \xi} - Mf_{i, \xi}^2) \right. \\ &\quad \left. - 2M \sum_{i=1}^{p-1} \sum_{j=i+1}^p \left(f_{i, \xi} \int ff_i \right) \left(f_{j, \xi} \int ff_j \right) \right] / (1 + MA_{p, \xi}). \end{aligned}$$

This expression can be simplified by a suitable combination of terms to give

$$S_p = \frac{-M \left(\int fA_p - f_{\xi} \right)^2}{1 + MA_{p, \xi}} + \sum_{i=1}^p \left(\int ff_i \right)^2 + Mf_{\xi}^2.$$

If we now take the limit of both sides of this equation and apply the Lemma and the hypothesis on the system $\{f_n\}$ we obtain the required result:

$$\begin{aligned} \lim_{n \rightarrow \infty} S_p &= \sum_{n=1}^{\infty} \pm \left(\int^* f \phi_n \right)^2 = \sum_{i=1}^{\infty} \left(\int ff_i \right)^2 + Mf_{\xi}^2 \\ &= \int f^2 + Mf_{\xi}^2 = \int^* f^2. \end{aligned}$$

THEOREM III. *If the systems $\{\phi_n\}$ and $\{f_n\}$ are M -complete and complete respectively, then $\{\phi_n\}$ is M -closed.*

To prove this theorem we assume the existence of a continuous function $\psi(x)$ which is M -orthogonal to all the ϕ_n . It is accordingly M -orthogonal to all the f_n , i. e.

$$\int^* \psi f_n = \int \psi f_n + M\psi_{\xi} f_{n, \xi} = 0, \quad (n = 1, 2, \dots).$$

If either or both of M and ψ_{ξ} are zero then $\psi \equiv 0$. If not, we have

$$\sum_{n=1}^{\infty} \left(\int \psi f_n \right)^2 = M^2 \psi_{\xi}^2 \sum_{n=1}^{\infty} f_{n, \xi}^2.$$

The series on the left is convergent, that on the right divergent by Theorem II. Hence the assumption is incorrect and $\{\phi_n\}$ is M -closed.

4. *The case of zero denominators.* The investigations so far have proceeded on the assumption that none of the denominators of (1a), or, what comes to the same thing, none of the factors P_n of (1c), is zero. Obviously P_n may be zero for one or more values of n . We now show that this contingency may be overcome by redefining the functions ϕ_n in terms of the f_n .

We first replace all conditions made hitherto on these f_n by the following:

- (a) They are continuous on I .
- (b) They are orthonormed on I .
- (c) The system $\{f_n\}$ is complete.
- (d) $\lim_{n \rightarrow \infty} A_{n,\xi} = \infty$.

If the first denominator in (1c) to vanish is that of ϕ_m , then

$$1 + MA_{m,\xi} = 0.$$

Thus $f_{m,\xi} \neq 0$, but any or all of $f_{i,\xi}$ ($1 \leq i \leq m-1$) may be zero. Also by condition (d) above, there exists a function f_i ($i > m$) of the set $\{f_n\}$ such that $f_{i,\xi} \neq 0$. Let f_{m+q} ($q \geq 1$) be the first such function of the set, i. e.

$$f_{i,\xi} = 0 \quad (m < i < m+q), \quad f_{(m+q),\xi} \neq 0.$$

Then the set of functions ϕ_n ($1 \leq n < m$) whose values are given by (1c), § 2, and the functions

$$\phi_n = f_{n+1} \quad (m \leq n \leq m+q-2)$$

are M -orthonormed on I . For each ϕ_n of this set P_n is positive.

If we express ϕ_{m+q-1} and ϕ_{m+q} each as linear combinations of the $m+q$ functions f_i ($1 \leq i \leq m+q$), and impose on them the conditions of M -orthonormality, we arrive at the expressions

$$\begin{aligned} \phi_{m+q-1} &= r \pm s \\ \phi_{m+q} &= r \mp s \end{aligned}$$

where

$$\begin{aligned} r &= [(f^2_{m,\xi} + f^2_{(m+q),\xi}) \sum_{i=1}^{m-1} f_{i,\xi} f_i + (f^2_{m,\xi} - f^2_{(m+q),\xi}) c f_{m,\xi} f_m \\ &\quad + 2c^2 f^2_{m,\xi} f_{(m+q),\xi} f_{m+q}] / 2c f^2_{m,\xi} f_{(m+q),\xi} \\ s &= [f^2_{m,\xi} \sum_{i=1}^{m-1} f_{i,\xi} f_i + f^3_{m,\xi} f_m] / 2c f^2_{m,\xi} f_{(m+q),\xi}. \end{aligned}$$

The constant c may not have the value zero, but otherwise it is arbitrary; for simplicity it will be taken equal to unity. Also

$$\int^* \phi^2_{m+q-1} = - \int^* \phi^2_{m+q} = \pm 1.$$

If we now take for ϕ_n ($n > m + q$) the values given by (1c), § 2, the system $\{\phi_n\}$ thus constructed is M -orthonormed on I , and can have no zero denominators.

Moreover $\int^* \phi_n^2 = +1$ for all values of n except one, either $n = m + q - 1$ or $n = m + q$.

Theorems I, II and III are valid for the system $\{\phi_n\}$ just obtained. We present here the proof of Theorem II only, as the others offer no difficulty.

If S_n represents the sum of the first n terms of the series

$$(3) \quad \sum_{i=1}^{\infty} \pm \left(\int^* f \phi_i \right)^2$$

then

$$\begin{aligned} S_{m+q+p} - S_{m+q} &= \sum_{i=m+q+1}^{m+q+p} f^2_{i,\xi} \left(\int f A_{m+q} - f_{\xi} \right)^2 / f^2_{(m+q),\xi} \sum_{i=m+q}^{m+q+p} f^2_{i,\xi} \\ &\quad + \sum_{i=m+q+1}^{m+q+p} \left(\int f f_i \right)^2 + I_p, \quad p \geq 1, \end{aligned}$$

where

$$I_p = \left[2f_{\xi} \sum_{i=m+q+1}^{m+q+p} f_{i,\xi} \int f f_i - \left(\sum_{i=m+q+1}^{m+q+p} f_{i,\xi} \int f f_i \right)^2 \right] / \sum_{i=m+q}^{m+q+p} f^2_{i,\xi}.$$

Since $\lim_{p \rightarrow \infty} I_p = 0$ we have

$$\lim_{p \rightarrow \infty} (S_{m+q+p} - S_{m+q}) = \left(\int f A_{m+q} - f_{\xi} \right)^2 / f^2_{(m+q),\xi} + \sum_{i=m+q+1}^{\infty} \left(\int f f_i \right)^2.$$

The sum of the first $m + q$ terms of the series (3) is given by

$$\begin{aligned} S_{m+q} &= \left[\int f A_{m-1} - f_{\xi} \right]^2 / f^2_{m,\xi} + \sum_{i=1}^{m-1} \left(\int f f_i \right)^2 + M f_{\xi}^2 \\ &\quad + \sum_{i=m+1}^{m+q-1} \left(\int f f_i \right)^2 + \left[(f^2_{m,\xi} + f^2_{(m+q),\xi}) \left(\int f A_{m-1} - f_{\xi} \right)^2 \right. \\ &\quad \left. - 2f^2_{m,\xi} \left(f_{m,\xi} \int f f_m + f_{(m+q),\xi} \int f f_{m+q} \right) \left(\int f A_{m-1} - f_{\xi} \right) \right. \\ &\quad \left. - f^2_{m,\xi} (f^2_{m,\xi} - f^2_{(m+q),\xi}) \left(\int f f_m \right)^2 \right. \\ &\quad \left. - 2f^3_{m,\xi} f_{(m+q),\xi} \int f f_m \int f f_{m+q} \right] / f^2_{m,\xi} f^2_{(m+q),\xi}. \end{aligned}$$

⁷ Cf. proof of Lemma, § 3.

Combining these results, we have after cancellation

$$\sum_{n=1}^{\infty} \pm \left(\int^* f \phi_n \right)^2 = \int f^2 + M f \xi^2 = \int^* f^2.$$

We are now justified in stating that if a set of functions $\{f_n\}$ exists which satisfies the four conditions enumerated above, then it is always possible to construct from it an M -complete, M -closed system $\{\phi_n\}$ of continuous functions which are linear combinations of the f_n , of the form

$$\phi_n = \sum_{i=1}^{p_n} c_{ni} f_i.$$

An example of such a set $\{f_n\}$ is given by the sequence of trigonometric functions

$$\{f_n\} = 1, 2^{1/2} \cos 2\pi x, 2^{1/2} \sin 2\pi x, 2^{1/2} \cos 4\pi x, 2^{1/2} \sin 4\pi x, \dots$$

In concluding this section we assemble the salient facts obtained so far in the

THEOREM IV. *For every M there exist M -complete, M -closed systems $\{\phi_n\}$ of functions which are continuous and M -orthogonal on I , and such that $\int^* \phi_n^2 = \pm 1$, where the negative sign occurs for one and only one value of n .*

SECTION II. EQUATION WITH SYMMETRIC KERNEL.

1. *Preliminary definitions and propositions.* We turn now to the equation

$$(1) \quad \lambda u(x) = \int^* K(x, s) u(s) ds,$$

in which the kernel $K(x, s)$ is a continuous symmetric^{*} function on R .

By a *characteristic number* of equation (1) is understood a value of the parameter λ for which the equation has a continuous solution $u(x)$ not identically zero on I .

In a similar fashion we employ the ordinary definitions of *characteristic function*, *pole of the resolvent*, *order of a pole*, etc., with the understanding that they refer to equation (1) unless definitely stated otherwise.

LEMMA. *If $K(x, s)$ is symmetric on R , then*

$$\int^* \int^* f(x) K(x, s) g(s) ds dx = \int^* \int^* f(s) K(s, x) g(x) ds dx$$

for all continuous functions f and g , and conversely.

^{*} $K(x, s)$ is symmetric on R if $K(x, s) \equiv K(s, x)$ on R , otherwise it is non-symmetric.

The first statement in the Lemma is readily verified. To prove the converse we assume that $K(x, s)$ is not symmetric, and arrive at a contradiction.

Application of the hypothesis to the equality gives

$$\int^* f(x) \left[\int^* \{K(x, s) - K(s, x)\} g(s) ds \right] dx = 0.$$

Let $\{\phi_n(x)\}$ be an M -closed system of continuous functions. If we take $f(x)$ equal, sequentially, to these functions ϕ_n , then

$$\int^* \phi_n(x) \left[\int^* \{K(x, s) - K(s, x)\} g(s) ds \right] dx = 0, \quad (n = 1, 2, \dots).$$

This is impossible, in view of the closure of the system $\{\phi_n\}$, unless

$$\int^* \{K(x, s) - K(s, x)\} g(s) ds \equiv 0.$$

If we take $g(s)$ equal, sequentially, to the same functions $\phi_n(s)$, and regard x as a parameter, then a similar process of reasoning gives

$$K(x, s) - K(s, x) \equiv 0 \text{ on } R,$$

i. e., $K(x, s)$ is symmetric on R .

2. Properties of characteristic numbers.

THEOREM I. *If the kernel $K(x, s)$ of (1) is symmetric on R , and if $u_1(x)$, $u_2(x)$ are characteristic functions of $K(x, s)$ for λ_1 , λ_2 respectively, ($\lambda_1 \neq \lambda_2$), then $u_1(x)$ and $u_2(x)$ are M -orthogonal on I .*

For, following Poisson's method of proof for the ordinary integral equation, we obtain

$$\lambda_1 \int^* u_1(x) u_2(x) dx = \int^* \int^* u_2(x) K(x, s) u_1(s) ds dx$$

$$\lambda_2 \int^* u_1(x) u_2(x) dx = \int^* \int^* u_1(x) K(x, s) u_2(s) ds dx,$$

hence, if we subtract these and apply the preceding lemma we get

$$\lambda_1 \int^* u_1(x) u_2(x) dx = \lambda_2 \int^* u_1(x) u_2(x) dx, \quad \lambda_1 \neq \lambda_2,$$

and from this follows

$$\int^* u_1(x) u_2(x) dx = 0.$$

THEOREM II. *A sufficient condition that all of the characteristic numbers of equation (1) for a symmetric kernel $K(x, s)$ be real is that $K(x, s)$ satisfy the inequality*

$$\int^* \int^* f(x) K(x, s) f(s) ds dx < 0 \text{ or } > 0^*$$

for all continuous functions $f \not\equiv 0$ on I .

Assume that $u(x) = v(x) + iw(x)$ is a complex characteristic function corresponding to $\lambda \neq 0$. Multiply (1) throughout by the conjugate characteristic function $\bar{u}(x) = v(x) - iw(x)$ and M -integrate with respect to x . We get

$$\begin{aligned} \lambda \int^* \bar{u}(x) u(x) dx &= \int^* \int^* \bar{u}(x) K(x, s) u(s) ds dx \\ &= \int \int v(x) K(x, s) v(s) ds dx + \int \int w(x) K(x, s) w(s) ds dx \\ &\quad + 2M[v(\xi) \int K(x, \xi) v(x) dx + w(\xi) \int K(x, \xi) w(x) dx] \\ &\quad + M^2 K(\xi, \xi) [v^2(\xi) + w^2(\xi)]. \end{aligned}$$

Since $\int^* \bar{u}(x) u(x) dx$ is real, and the right-hand side is real, a sufficient condition that λ be real is that $\int^* \bar{u}u \neq 0$, or that the expression on the right be different from zero. This is satisfied if

$$\int^* \int^* f K f > 0 \text{ or } < 0$$

for all continuous functions $f \not\equiv 0$ on I , which was to be proved.

COROLLARY I. *If in the pure integral equation*

$$(2) \quad \mu \phi(x) = \int K(x, s) \phi(s) ds$$

* For example, if the system $\{\phi_n\}$ is M -closed on I , and if

$$K(x, s) = \sum_{n=1}^{\infty} \frac{\phi_n(x) \phi_n(s)}{\eta_n^2},$$

the η_n^2 so chosen that the series converges uniformly on I , then

$$\int^* \int^* f K f = \sum_{n=1}^{\infty} \frac{1}{\eta_n^2} \left(\int^* f(x) \phi_n(x) dx \right)^2 \geq 0,$$

the equality sign holding only if $f \equiv 0$.

the kernel $K(x, s)$ is positive,¹⁰ and if the characteristic functions ϕ_a of (2) form a complete orthonormed system for this $K(x, s)$, and are such that

$$\lim_{n \rightarrow \infty} \sum_{a=1}^n \phi_a^2 = \infty,$$

then the characteristic numbers of (1) for the same kernel are real.

The proof is based upon two theorems of Mercer's¹⁰ by which we may write

$$K(x, s) = \sum_{a=1}^{\infty} \mu_a \phi_a(x) \phi_a(s),$$

the characteristic numbers μ_a of (2) being all positive. Then

$$\int^* \int^* f K f = \sum_{a=1}^{\infty} \mu_a \left(\int^* f \phi_a \right)^2.$$

The series on the right may be zero only if $f \equiv 0$, since otherwise

$$\int^* f \phi_a = \int f \phi_a + M f_{\xi} \phi_a, \xi = 0, \quad (\alpha = 1, 2, \dots),$$

$$\therefore \sum_{a=1}^{\infty} \left(\int f \phi_a \right)^2 = M^2 f_{\xi}^2 \sum_{a=1}^{\infty} \phi_a^2;$$

the series on the left of the last equation is convergent, that on the right divergent, hence

$$\int^* \int^* f K f \geq 0,$$

the equality holding only if $f \equiv 0$; therefore the characteristic numbers of (1) are real by the previous theorem.

COROLLARY 2. *A sufficient condition that all of the characteristic numbers of equation (1) for a symmetric kernel $K(x, s)$ be real is that $K(x, s)$ satisfy the inequality*

¹⁰ We use here Mercer's definition of a positive function: A symmetric function $K(x, s)$ is positive if

$$\int \int f K f \geq 0$$

for all continuous functions f . See *Phil. Trans. Royal Soc. A*, vol. 209 (1909), pp. 417-444.

$$\iint [K(\xi, x)K(\xi, s) - K(\xi, \xi)K(x, s)]f(x)f(s)ds dx < 0$$

for all continuous functions $f \not\equiv 0$.¹¹

For the inequality of Theorem II is true for all $f(\xi)$ if its discriminant is negative, i. e., if

$$M^2 \left[\int K(x, \xi)f(x)dx \right]^2 - K(\xi, \xi)M^2 \iint f(x)K(x, s)f(s)ds dx < 0,$$

hence if

$$\iint [K(x, \xi)K(\xi, s) - K(\xi, \xi)K(x, s)]f(x)f(s)ds dx < 0$$

for all continuous $f \not\equiv 0$.

THEOREM III. EXISTENCE THEOREM. *Every symmetric kernel has at least one characteristic number.*

To facilitate the proof we first introduce the following definition:

A matrix $A = (a_{ij})$ is *M-symmetric* if it is symmetric except for one row, say the N -th, and its corresponding column, and if the elements a_{Ni} and a_{iN} of this row and column satisfy the relations

$$a_{Ni} = -a_{iN}, \quad i \neq N.$$

As was stated in the Introduction we use Hilbert's method¹² for passing from an integral equation to a system of linear equations in infinitely many unknowns.

Let $\{\phi_n\}$ be an M -complete system of the type defined by formula (1c), Section I. Multiply equation (1) throughout by $\phi_i(x)$, M -integrate with respect to x , and apply Theorem II,¹³ Section I; this gives

¹¹ An example of such a kernel is

$$K(x, s) = 1 + \sum_{n=1}^{\infty} \frac{\sin n(x-s) \cdot \sin n(x-\xi)}{\eta_n^2},$$

since

$$\begin{aligned} & - \iint [K(\xi, x)K(\xi, s) - K(\xi, \xi)K(x, s)]f(x)f(s)ds dx \\ & = - \sum_{n=1}^{\infty} \frac{1}{\eta_n^2} \left(\int f(x) \sin n(x-\xi) \right)^2 < 0 \text{ for } f \not\equiv 0. \end{aligned}$$

¹² Hilbert, *Grundzüge einer allgemeinen Theorie der linearen integral-Gleichungen*, pp. 186-188.

¹³ In Section I we assumed implicitly that the arbitrary function f was real. There is nothing in the proof of Theorem II, Section I, which will not be valid if f is a continuous complex function.

$$\begin{aligned}\lambda \int^* \phi_i u &= \int^* \int^* \phi_i K u \\ &= \sum_{K=1}^{\infty} \pm \int^* \int^* \phi_i K \phi_K \int^* u \phi_K, \quad (i = 1, 2, \dots).\end{aligned}$$

If we introduce the substitutions

$$\int^* \phi_i u = x_i, \quad \pm \int^* \int^* \phi_i K \phi_K = a_{iK}$$

the above equations are replaced by the linear system

$$(3) \quad \lambda x_i = \sum_{K=1}^{\infty} a_{iK} x_K, \quad (i = 1, 2, \dots),$$

in the infinitely many unknowns x_i .

If we let the x_i represent the components of a vector x in a space of infinitely many dimensions, then x is a vector of the Hilbert space,¹⁴ for

$$\sum_{i=1}^{\infty} \pm x_i \bar{x}_i = \sum_{i=1}^{\infty} \pm \int^* u \phi_i \int^* \bar{u} \phi_i = \int^* u \bar{u},$$

hence

$$\sum_{i=1}^{\infty} x_i \bar{x}_i \text{ converges.}$$

The matrix of (3) is M -symmetric, and it is completely continuous.¹⁵

To prove the latter property it is sufficient to demonstrate that $\sum_{iK=1}^{\infty} a^2_{iK}$ is convergent.

By definition

$$\sum_{iK=1}^{\infty} a^2_{iK} = \sum_{iK=1}^{\infty} \left[\int^* \int^* \phi_i K \phi_K \right]^2.$$

If Theorem II, Section I is applied twice to the series on the right of this equality we obtain successively

$$\begin{aligned}\sum_{K=1}^{\infty} \left[\int^* \int^* \phi_i K \phi_K \right]^2 &= \int^* \left[\int^* \phi_i K \right]^2 + 2 \left[\int^* \int^* \phi_i K \phi_N \right]^2 \\ \sum_{iK=1}^{\infty} \left[\int^* \int^* \phi_i K \phi_K \right]^2 &= \sum_{i=1}^{\infty} \int^* \left[\int^* \phi_i K \right]^2 + 2 \left[\int^* K \phi_N \right]^2 \\ &\quad + 4 \left[\int^* \int^* \phi_N K \phi_N \right]^2.\end{aligned}$$

¹⁴ A vector $x = (x_i)$ is said to belong to the Hilbert space if $\sum_{i=1}^{\infty} x_i \bar{x}_i$ converges.

¹⁵ Hilbert, *Grundzüge*, Kap. XI, p. 147 f. The convergence of $\sum_{iK=1}^{\infty} a^2_{iK}$ is a necessary and sufficient condition for the complete continuity of the matrix A .

The series $\sum_{i=1}^{\infty} \left[\int^* \phi_i K \right]^2$ is uniformly convergent, since it converges to the continuous function

$$\int^* K^2 + 2 \left[\int^* K \phi_N \right]^2$$

and since the terms of the series are positive.¹⁶

Application of Theorem II, § 3 of Section I to this series yields

$$\sum_{i=1}^{\infty} \int^* \left[\int^* \phi_i K \right]^2 = \int^* \sum_{i=1}^{\infty} \left[\int^* \phi_i K \right]^2 = \int^* \int^* K^2 + 2 \int^* \left[\int^* K \phi_N \right]^2,$$

hence $\sum_{iK=1}^{\infty} a^2_{iK}$ is convergent.

It has been proved¹⁷ that a system of equations of the type (3), the matrix of which is M -symmetric, has at least one characteristic number. Let x then, be a characteristic vector of (3) belonging to the Hilbert space. The series

$$(4) \quad \sum_{K=1}^{\infty} \pm x_K \int^* K \phi_K$$

is absolutely and uniformly convergent on I . This follows at once if we apply the Lagrange-Cauchy inequality to the series obtained from (4) by omitting those terms, at most two in number,¹⁸ which correspond to complex values of λ . Denote the sum of the series by $\lambda u(x)$:

$$\lambda u(x) = \sum_{K=1}^{\infty} \pm x_K \int^* K \phi_K,$$

multiply this equation throughout by $\phi_i(x)$, and M -integrate with respect to x ; then

$$\lambda \int^* \phi_i u = \sum_{K=1}^{\infty} \pm x_K \int^* \int^* \phi_i K \phi_K.$$

Since, by hypothesis, x is a characteristic vector of the system of linear equations (3) we have

$$x_i = \int^* \phi_i u,$$

and therefore

$$\begin{aligned} \lambda u(x) &= \sum_{K=1}^{\infty} \pm \int^* u \phi_K \int^* K(x, s) \phi_K(s) ds \\ &= \int^* K(x, s) u(s) ds. \end{aligned}$$

¹⁶ E. W. Hobson, *The Theory of Functions of a Real Variable*, Sec. Ed., vol. 2, p. 116.

¹⁷ A. Pell-Wheeler, *loc. cit.*, Theorem II.

¹⁸ Cf. p. 878.

Thus a unique correspondence has been established between characteristic vectors of (3) and characteristic functions of the integral equation (1). The symmetric kernel $K(x, s)$ of (1) has therefore at least one characteristic number, which was to be proved.

In the Introduction examples were given of kernels with zero and imaginary characteristic numbers, and of characteristic numbers that are not simple poles of the resolvent. There are however, certain restrictions that can be made on them.

It can easily be proved¹⁹ that one characteristic number may be zero, that two may be imaginary (conjugate), the rest being real, that a characteristic number can not be a pole of the resolvent of order greater than three, and that there can not be more than one which is not a simple pole of the resolvent. A necessary condition that λ be not a simple pole of the resolvent of equation (1) is that $\int^* u^2 = 0$.²⁰

3. *Expansion theorem.* Our problem is this: to determine sufficient conditions upon a continuous function $g(x)$ in order that it may be expressible in the form

$$g(x) = \sum_a c_a u_a(x),$$

where the u_a are the characteristic functions of the integral equation (1) for the symmetric kernel $K(x, s)$.

Certain functions $U_a(x)$ ²¹ will be introduced corresponding to a characteristic number λ_m which is not a simple pole of the resolvent. These functions satisfy the relations

$$\begin{aligned} \int^* U_\alpha U_\beta &= \pm e_{\alpha\beta} \\ \int^* K U_\alpha &= \sum_\beta \lambda_m c_{\alpha\beta} U_\beta, \quad |c_{\alpha\beta}| \neq 0, \quad \lambda_m \neq 0, \\ \int^* K U_\alpha &= \sum_\beta c_{\alpha\beta} U_\beta, \quad |c_{\alpha\beta}| = 0, \quad \lambda_m = 0. \end{aligned}$$

A. Pell-Wheeler has proved¹⁹ that the iterated matrix $A^{(4)} = (a_{iK}^{(4)})$ is developable into the following series:

$$(5) \quad a_{iK}^{(4)} = \sum_{j=1}^4 E_j^{(4)},$$

¹⁹ A. Pell-Wheeler, *loc. cit.*, Theorem I.

²⁰ That the condition is not sufficient is illustrated by Example 1, Introduction.

²¹ These functions U_a are analogous to the so-called "principal functions" of the Fredholm theory.

where $E_1^{(4)}$ is that part of the expansion corresponding to real characteristic numbers λ_α which are simple poles of the resolvent, and is an absolutely convergent series; $E_2^{(4)}$ contains those terms corresponding to imaginary λ ; $E_3^{(4)}$ those terms corresponding to $\lambda \neq 0$ not a simple pole of the resolvent; and $E_4^{(4)}$ those terms corresponding to $\lambda = 0$ not a simple pole of the resolvent.

The expressions written out in full for $E_j^{(4)}$ are

$$E_1^{(4)} = \sum_a \pm \lambda_a^4 s_{ai} s_{aK}^*,$$

the s_a being characteristic vectors for the matrix A , and also $A^{(4)}$, and satisfying the relations

$$(s_a s_\beta^*) = \sum_i s_{ai} s_{\beta i}^* = \pm e_{a\beta};$$

$$E_2^{(4)} = \pm (\lambda^4 s_i s_K^* + \lambda^4 \bar{s}_i \bar{s}_K^*),$$

where

$$(s \bar{s}^*) = 0, \quad (s s^*) = \pm 1;$$

$$E_3^{(4)} = \lambda_m^4 \sum_{\alpha\beta} \pm c_{\alpha\beta} s_{ai} s_{\beta K}^*, \quad |c_{\alpha\beta}| \neq 0, \quad (\alpha, \beta = 1, 2 \text{ or } 1, 2, 3),$$

where

$$(s_a s_\beta^*) = \pm e_{a\beta}, \quad \sum_K a_{iK}^{(4)} s_{aK} = \lambda_m^4 \sum_\beta c_{a\beta} s_{\beta K};$$

$$E_4^{(4)} = \sum_{\alpha\beta} \pm c_{\alpha\beta} s_{ai} s_{\beta K}^*, \quad |c_{\alpha\beta}| = 0, \quad (\alpha, \beta = 1, 2 \text{ or } 1, 2, 3),$$

where

$$(s_a s_\beta^*) = \pm e_{a\beta}, \quad \sum_K a_{iK}^{(4)} s_{aK} = \sum_\beta c_{a\beta} s_{\beta K}.$$

From (5) we have

$$(6) \quad \sum_{iK} a_{iK}^{(4)} x_i y_K = \sum_{iK} \left[\sum_j E_j \right] x_i y_K,$$

the series being absolutely and uniformly convergent for all vectors x and y of the Hilbert space. This condition is satisfied if we take

$$x_i = \int^* f \phi_i \quad \text{and} \quad y_i = \int^* g \phi_i,$$

the functions f and g being continuous on I . Substitute these values in (6), and the expression on the left becomes

$$\begin{aligned} \sum_{iK} a_{iK}^{(4)} x_i y_K &= \sum_{iK} \pm \int^* \int^* \phi_i K^{(4)} \phi_K \int^* f \phi_i \int^* g \phi_K \\ &= \int^* \int^* K^{(4)} f g, \end{aligned}$$

by Theorem II, Section I.

If we make the same substitutions in the right side of (6) we obtain

$$\begin{aligned}\sum_{iK} E_1^{(4)} x_i y_K &= \sum_a \pm \lambda_a^4 \sum_i \pm \int^* u_a \phi_i \int^* f \phi_i \sum_K \pm \int^* u_a \phi_K \int^* g \phi_K \\ &= \sum_a \pm \lambda_a^4 \int^* u_a f \int^* u_a g,\end{aligned}$$

the series being absolutely and uniformly convergent;

$$\begin{aligned}\sum_{iK} E_2^{(4)} x_i y_K &= \pm \left[\lambda^4 \sum_i \pm \int^* u \phi_i \int^* f \phi_i \sum_K \int^* u \phi_K \int^* g \phi_K \right. \\ &\quad \left. + \bar{\lambda}^4 \sum_i \pm \int^* \bar{u} \phi_i \int^* f \phi_i \sum_K \int^* \bar{u} \phi_K \int^* g \phi_K \right] \\ &= \pm \left(\lambda^4 \int^* u f \int^* u g + \bar{\lambda}^4 \int^* \bar{u} f \int^* \bar{u} g \right);\end{aligned}$$

and the remaining terms of (6), corresponding to a characteristic number not a simple pole of the resolvent, are respectively

$$\begin{aligned}\sum_{iK} E_3^{(4)} x_i y_K &= \lambda_m^4 \sum_{a\beta} \pm c_{a\beta} \sum_i s_{ai} x_i \sum_K s_{\beta K}^* y_K, & |c_{a\beta}| \neq 0, \\ \sum_{iK} E_4^{(4)} x_i y_K &= \sum_{a\beta} \pm c_{a\beta} \sum_i s_{ai} x_i \sum_K s_{\beta K}^* y_K, & |c_{a\beta}| = 0.\end{aligned}$$

The vectors s_a appearing in these two expressions are equal to the same linear combinations of characteristic vectors for λ_m as the functions U_a , introduced above, are of the characteristic functions of (1), corresponding to the same λ_m ; and the conditions imposed upon the s_a lead, under the transformation

$$s_{ai} = \int^* U_a \phi_i$$

to the conditions imposed upon the U_a . The converse is also true. Hence

$$\begin{aligned}\sum_{iK} E_3^{(4)} x_i y_K &= \lambda_m^4 \sum_{a\beta} \pm c_{a\beta} \int^* U_a f \int^* U_\beta g, & |c_{a\beta}| \neq 0, \\ \sum_{iK} E_4^{(4)} x_i y_K &= \sum_{a\beta} \pm c_{a\beta} \int^* U_a f \int^* U_\beta g, & |c_{a\beta}| = 0.\end{aligned}$$

Combining these results, we have

$$\begin{aligned}(7) \quad \int^* \int^* K^{(4)} fg &= \sum_a \pm \lambda_a^4 \int^* u_a f \int^* u_a g \\ &\quad \pm \left(\lambda^4 \int^* u f \int^* u g + \bar{\lambda}^4 \int^* \bar{u} f \int^* \bar{u} g \right) \\ &\quad + \lambda_m^4 \sum_{a\beta} \pm c_{a\beta} \int^* U_a f \int^* U_\beta g \\ &\quad + \sum_{a\beta} \pm c_{a\beta} \int^* U_a f \int^* U_\beta g.\end{aligned}$$

Since the series

$$\sum_a \pm \lambda_a^4 \left(\int^* u_a v \right)^2$$

converges for all continuous functions $v(x)$, then the series

$$\sum_a \pm \lambda_a^6 u_a^2(x),$$

and hence the series

$$\sum_a \pm \lambda_a^6 u_a \int^* u_a v$$

converges absolutely and uniformly.

Set

$$\begin{aligned} (8) \quad h(x) = & \int^* K^{(5)} f - \left[\sum_a \lambda_a^4 u_a \int^* \int^* f K^{(4)} u_a \pm \left(\lambda^4 u \int^* \int^* f K^{(4)} u \right. \right. \\ & + \bar{\lambda}^4 \bar{u} \int^* \int^* f K^{(4)} \bar{u} \left. \right) + \sum_{\alpha\beta} \pm \lambda_m^4 c_{\alpha\beta} U_\alpha \int^* \int^* f K^{(4)} U_\beta \\ & \left. + \sum_{\alpha\beta} \pm c_{\alpha\beta} U_\alpha \int^* \int^* f K^{(4)} U_\beta \right]. \end{aligned}$$

If l is an arbitrary continuous function, then

$$\begin{aligned} \int^* h l = & \int^* \int^* l K^{(5)} f - \left[\sum_a \lambda_a^4 \int^* l u_a \int^* \int^* f K^{(4)} u_a \right. \\ & \pm \left(\lambda^4 \int^* l u \int^* \int^* f K^{(4)} u + \bar{\lambda}^4 \int^* l \bar{u} \int^* \int^* f K^{(4)} \bar{u} \right) \\ & + \sum_{\alpha\beta} \pm \lambda_m^4 c_{\alpha\beta} \int^* l U_\alpha \int^* \int^* f K^{(4)} U_\beta \\ & \left. + \sum_{\alpha\beta} \pm c_{\alpha\beta} \int^* l U_\alpha \int^* \int^* f K^{(4)} U_\beta \right]. \end{aligned}$$

Since the right side of this equality is zero from (7), it follows that $h \equiv 0$ by the Lemma, Section II.

If we now choose g such that

$$g(x) = \int^* K^{(5)} f$$

and make the substitution in (8), we have an expansion of g in the form

$$\begin{aligned} g(x) = & \sum_a \pm u_a \int^* g u_a \pm \left(u_a \int^* g u_a + \bar{u}_a \int^* g \bar{u}_a \right) \\ & + \sum_a \pm U_a \int^* g U_a. \end{aligned}$$

In conclusion we set forth the

THEOREM IV. *Every function of the form*

$$g(x) = \int^* K^{(5)}(x, s)f(s)ds,$$

where $f(s)$ is an arbitrary function continuous on I , can be expressed in the form of an absolutely and uniformly convergent series:

$$g(x) = \sum_a \pm u_a(x) \int^* g u_a + E(x),$$

where the u_a are characteristic functions of the integral equation (1), and $E(x)$ consists of a finite number of terms arising from the existence of characteristic numbers which are conjugate imaginary, or zero, or are poles of the resolvent of order greater than one.

BRYN MAWR COLLEGE.

ON THE EXPANSION OF HARMONIC FUNCTIONS IN SERIES OF HARMONIC POLYNOMIALS BELONGING TO A SIMPLY CONNECTED REGION.¹

By O. J. FARRELL.

1. *Introduction.* In his address ² to the American Mathematical Society on the approximation of harmonic functions by harmonic polynomials and by harmonic rational functions Professor Walsh included a treatment of expansions in given regions in terms of a particular set of harmonic polynomials belonging to that region. Among the results reported in this connection was a theorem of his own to the effect that if C be a simple finite analytic curve in the (x, y) -plane, then there exist harmonic polynomials $\{p_n(x, y)\}$ such that if $f(x, y)$ is defined and continuous on C and on C is of bounded variation then $f(x, y)$ can be developed into a series

$$(a) \quad f(x, y) = \sum_{k=1}^{\infty} a_k p_k(x, y)$$

which converges uniformly in the closed interior of C . Series (a) thus represents a function harmonic interior to C , continuous in the corresponding closed region, and having the value $f(x, y)$ on C . There exist continuous functions $\{q_n(x, y)\}$ on C with which the polynomials $\{p_n(x, y)\}$ form a biorthogonal set:

$$\int_C p_k(x, y) q_m(x, y) ds = \begin{cases} 0, & \text{if } k \neq m \\ 1, & \text{if } k = m. \end{cases}$$

The coefficients of (a) are given by the formulas

$$a_k = \int_C f(x, y) q_k(x, y) ds.$$

The functions $\{q_n(x, y)\}$ depend on C but not on $f(x, y)$.

If the curve C of this theorem is the unit circle, the situation is classical; and if polar coördinates (ρ, ϕ) be introduced in the (x, y) -plane, the functions $\{p_n(x, y)\}$ may be chosen directly as the set of harmonic polynomials $\{\rho^n \cos n\phi, \rho^n \sin n\phi\}$. For the Fourier development of $f(x, y)$ on the unit

¹ Presented to the American Mathematical Society, December 27, 1929 and September 7, 1934.

² Walsh, *Bulletin of the American Mathematical Society*, vol. 35 (1929), pp. 499-544.

circle gives rise at once to a series in terms of these functions, and this series converges uniformly throughout the closed interior of the unit circle.

The report of the theorem just quoted was followed by mention of the fact that the theorem would lend itself readily to generalizations. It is the purpose of the present paper to contribute one such generalization as follows:

THEOREM I. *In the (x, y) -plane let G be a limited simply connected region whose boundary B consists wholly of simple³ boundary points and is also the boundary of an infinite region. Let the region G be mapped conformally onto the interior of the unit circle in the w -plane, $w = x' + iy' = \rho e^{i\phi}$, by means of an analytic function $w = \Phi(z)$, $z = x + iy$, and denote by C_μ the transform in the (x, y) -plane of the circle $\rho = \mu$ in the (x', y') -plane. Then there exist harmonic polynomials $\{p_n(x, y)\}$ and a series of functions $\{q_n(x, y)\}$ continuous and biorthogonal on every curve C_μ , $0 < \mu < 1$, to the polynomials $\{p_n(x, y)\}$:*

$$\int_{C_\mu} p_k(x, y) q_m(x, y) ds = \begin{cases} 0, & \text{if } k \neq m \\ 1, & \text{if } k = m. \end{cases}$$

If $f(x, y)$ is an arbitrary function harmonic in G , and if C_μ be the transform of an arbitrarily chosen circle $\rho = \mu$, $0 < \mu < 1$, then the series

$$(1) \quad \sum_{k=1}^{\infty} a_k p_k(x, y), \quad a_k = \int_{C_\mu} f(x, y) q_k(x, y) ds,$$

converges to $f(x, y)$ continuously⁴ in G . If $f(x, y)$ is harmonic interior to C_μ , then the formal expansion (1) of $f(x, y)$ found by integration over C_μ , $0 < \mu' < \mu$, converges to $f(x, y)$ continuously in the region interior to C_μ . If $f(x, y)$ is merely known to be defined and continuous on C_μ , then the formal expansion (1) of $f(x, y)$ found by integration on C_μ is summable (C_1) to $f(x, y)$ uniformly on C_μ , hence summable (C_1) uniformly on and within C_μ , thus furnishing a solution of the Dirichlet problem for the region interior to C_μ and the boundary values $f(x, y)$.

If the region G of Theorem I happens to be the interior of the unit circle, the situation is again the classical one, and the functions $\{p_n(x, y)\}$ may be

³ A boundary point is simple if it is contained in just one primend. See Carathéodory, *Mathematische Annalen*, vol. 73 (1913), pp. 321-370, §§ 44-46.

⁴ A series is said to converge continuously in a region if in any closed subregion the series converges uniformly. See Walsh, *Transactions of the American Mathematical Society*, vol. 33 (1931), pp. 668-689, § 2.

taken as the harmonic polynomials $\{\rho^n \cos n\phi, \rho^n \sin n\phi\}$. The curves C_μ are now the circles $\rho = \mu$, $0 < \mu < 1$; and the biorthogonal q -functions are the functions $\{(1/\pi\rho^{n+1}) \cos n\phi, (1/\pi\rho^{n+1}) \sin n\phi\}$.

To obtain Theorem I we first consider (in Theorems II and III below) expansions of functions harmonic within the unit circle in the (x', y') -plane. These expansions are not in terms of harmonic polynomials but in terms of a set of harmonic functions $\{p'_n(x', y')\}$ which approximate to the functions $\{(1/\pi^{\frac{1}{2}})\rho^n \cos n\phi, (1/\pi^{\frac{1}{2}})\rho^n \sin n\phi\}$. If the approximations be taken sufficiently close, an arbitrary function $f'(x', y')$ harmonic within the unit circle can be expanded into a series in terms of the functions $\{p'_n(x', y')\}$ where the coefficients of the expansion are found with the help of a biorthogonal set of functions by integration over an arbitrary circle γ_μ : $\rho = \mu$, $0 < \mu < 1$. The region G of Theorem I is then mapped onto the interior of the unit circle in the (x', y') -plane of Theorem III and by means of the properties of the conformal map the result of Theorem III yields its analog in the (x, y) -plane and gives us Theorem I.

2. *Developments within the unit circle.* The point of departure here is the following specialization of a theorem of Walsh:⁵

THEOREM II. Let the functions $\{\pi_n(w)\}$, $w = x' + iy' = \rho e^{i\phi}$, be analytic for $|w| \leq 1 + \epsilon$, $\epsilon > 0$, and such that for $|w| \leq 1 + \epsilon$ we have

$$(2) \quad |\pi_n(w) - (1/\pi^{\frac{1}{2}})w^n| \leq \bar{\epsilon}_n, \quad (n = 1, 2, \dots),$$

where the series $\sum \bar{\epsilon}_n$ converges and the sum of the series $\sum \bar{\epsilon}_n^2$ is less than $1/4\pi$. Let

$$(3) \quad \begin{aligned} \pi_0(w) &= p'_1(x', y') = 1/(2\pi)^{\frac{1}{2}} \\ \pi_n(w) &= p'_{2n}(x', y') + ip'_{2n+1}(x', y'), \end{aligned} \quad (n = 1, 2, \dots).$$

Then there exists a set of functions $\{q'_n(x', y')\}$ defined and continuous on the unit circle γ : $\rho = 1$ and such that

$$\int_{\gamma} p'_k(x', y') q'_m(x', y') d\phi = \begin{cases} 0, & \text{if } k \neq m \\ 1, & \text{if } k = m. \end{cases}$$

Any function $f'(x', y')$ which is continuous and of bounded variation on γ can be developed into a series

$$(4) \quad f'(x', y') = \sum_{k=1}^{\infty} a_k p'_k(x', y'), \quad a_k = \int_{\gamma} f'(x', y') q'_k(x', y') d\phi,$$

⁵ *Proceedings of the National Academy of Sciences*, vol. 13 (1927), pp. 175-180, Theorem 3.

which converges uniformly throughout the closed interior of γ , thus defining a function harmonic for $\rho < 1$, continuous for $\rho \leq 1$, and equal to $f'(x', y')$ on γ . If $f'(x', y')$ is merely known to be continuous on γ , then the formal development (4) of $f'(x', y')$ on γ is summable (C_1) to $f'(x', y')$ uniformly on γ , hence summable (C_1) uniformly on and within γ , thus furnishing a solution of the Dirichlet problem for the region interior to γ and the boundary values $f'(x', y')$.

It is seen at once that Theorem II differs from the result as given by Walsh only in that the functions $\{p'_n(x', y')\}$ have been specialized. For if $\pi_n(w)$ is analytic in the closed interior of the circle $\rho = 1 + \epsilon$, the real and imaginary parts of $\pi_n(w)$ are harmonic in this closed region. And inequalities (2) imply the original inequalities (5) of Walsh's Theorem 3. For it follows from (2) that

$$\begin{aligned} |p'_{2n}(x', y') - (1/\pi^{\frac{1}{2}})\rho^n \cos n\phi| &\leq \bar{\epsilon}_n \\ |p'_{2n+1}(x', y') - (1/\pi^{\frac{1}{2}})\rho^n \sin n\phi| &\leq \bar{\epsilon}_n \quad (n = 1, 2, \dots). \end{aligned}$$

And if the $\bar{\epsilon}_n$ of (2) be identified with the ϵ_n required by Walsh by setting $\bar{\epsilon}_n = \epsilon_{2n} = \epsilon_{2n+1}$ ($n = 1, 2, \dots$), we have, since $\epsilon_1 = 0$,

$$\sum_{n=1}^{\infty} \epsilon_n^2 = \sum_{n=2}^{\infty} \epsilon_n^2 = 2 \sum_{n=1}^{\infty} \bar{\epsilon}_n^2 < 1/2\pi.$$

Theorem II will now be applied to prove the following theorem.

THEOREM III. For $|w| < 1$ let the functions $\pi_n(w)$ ($n = 1, 2, \dots$), be analytic and such that

$$(5) \quad |\pi_n(w) - (1/\pi^{\frac{1}{2}})w^n| \leq \lambda^n \bar{\epsilon}_n \quad (n = 1, 2, \dots),$$

where λ is any fixed positive number less than unity, and where the $\bar{\epsilon}_n$ satisfy the hypotheses of Theorem II. Suppose further that $\pi_n(w)$ has a zero of order n at the origin. Let the functions $\{p'_n(x', y')\}$ be defined as in (3) of Theorem II. Then there exists a set of functions $\{q''_m(x', y')\}$ which are continuous and biorthogonal to the set $\{p'_n(x', y')\}$ on every circle $\gamma_\mu: \rho = \mu$, $0 < \mu \leq 1$:

$$\int_{\gamma_\mu} p'_k(x', y') q''_m(x', y') d\sigma = \begin{cases} 0, & \text{if } k \neq m \\ 1, & \text{if } k = m, \end{cases}$$

where σ denotes arc length along γ_μ . If $f'(x', y')$ is any function harmonic within the unit circle γ , then the series

$$(6) \quad \sum_{k=1}^{\infty} a_k p'_k(x', y'), \quad a_k = \int_{\gamma_\mu} f'(x', y') q''_k(x', y') d\sigma,$$

converges to $f'(x', y')$ continuously thruout the interior of γ . If $f'(x', y')$ is harmonic interior to γ_μ , then the formal development (6) of $f'(x', y')$ found by integration over γ_μ , $0 < \mu' < \mu$, converges to $f'(x', y')$ continuously interior to γ_μ . If $f'(x', y')$ is merely known to be defined and continuous on γ_μ , then the formal expansion (6) of $f'(x', y')$ on γ_μ is summable (C_1) to $f'(x', y')$ uniformly on γ_μ hence is summable (C_1) uniformly on and within γ_μ , thus furnishing a solution of the Dirichlet problem for the region interior to γ_μ and the boundary values $f'(x', y')$.

To prove this theorem we shall show first that if μ be chosen arbitrarily between zero and unity, there is a positive ϵ' such that

$$(7) \quad |\pi_n(\mu w)/\mu^n - (1/\pi^{\frac{1}{2}})w^n| \leq \bar{\epsilon}_n, \quad |w| < 1 + \epsilon'.$$

From (5) it follows that

$$|\pi_n(w)/w^n - 1/\pi^{\frac{1}{2}}| \leq \lambda^n \bar{\epsilon}_n / |w|^n, \quad w \neq 0, \quad |w| < 1.$$

Moreover, since $\pi_n(w)$ has a zero of order n at the origin, the function $g_n(w) = \pi_n(w)/w^n - 1/\pi^{\frac{1}{2}}$ when properly defined at the origin is analytic everywhere within the unit circle. The inequalities (5) may be written

$$|w^n g_n(w)| / \lambda^n \bar{\epsilon}_n \leq 1, \quad |w| < 1,$$

whence by Schwarz's Lemma

$$|w^n g_n(w)| / \lambda^n \bar{\epsilon}_n \leq |w|, \quad |w| < 1,$$

that is

$$|w^{n-1} g_n(w)| / \lambda^n \bar{\epsilon}_n \leq 1, \quad |w| < 1.$$

By repeated use of Schwarz's Lemma we finally get

$$|g_n(w)| \leq \lambda^n \bar{\epsilon}_n, \quad |w| < 1.$$

Since this last inequality holds for all w of modulus less than unity, it will hold when w is replaced by μw , $0 < \mu < 1$. Accordingly we have

$$|\pi_n(\mu w)/\mu^n w^n - 1/\pi^{\frac{1}{2}}| \leq \lambda^n \bar{\epsilon}_n, \quad \mu |w| < 1,$$

or

$$|\pi_n(\mu w)/\mu^n - (1/\pi^{\frac{1}{2}})w^n| \leq \lambda^n \bar{\epsilon}_n |w|^n, \quad \mu |w| < 1.$$

The quantity $\lambda^n \bar{\epsilon}_n |w|^n$ is not greater than $\bar{\epsilon}_n$ if $|w| < 1/\lambda$. If then we choose a positive ϵ' such that $1 + \epsilon'$ is less than the smaller of $1/\lambda$ and $1/\mu$, we have the desired inequalities (7).

Let $f'(x', y')$ be a function harmonic within the unit circle. Choose a number μ between zero and unity. By Theorem II the function $f'(\mu x', \mu y')$ admits an expansion

$$(8) \quad f'(\mu x', \mu y') = b_1 p'_1(\mu x', \mu y') + \cdots + b_{2n} p'_{2n}(\mu x', \mu y') / \mu^n \\ + b_{2n+1} p'_{2n+1}(\mu x', \mu y') / \mu^n + \cdots,$$

which converges uniformly in the closed interior of γ . Moreover, there exists a set of functions $\{\bar{q}_n(x', y')\}$ continuous and biorthogonal on γ to the set

$$p'_1(\mu x', \mu y'), \cdots, p'_{2n}(\mu x', \mu y') / \mu^n, p'_{2n+1}(\mu x', \mu y') / \mu^n, \cdots,$$

and the coefficients of (8) are given by

$$b_k = \int_{\gamma} f'(\mu x', \mu y') \bar{q}_k(x', y') d\phi.$$

But (8) is equivalent to a development of $f'(x', y')$ converging uniformly on and within γ_{μ} , namely,

$$(9) \quad f'(x', y') = \sum_{k=1}^{\infty} a_k p'_k(x', y'), \quad a_k = \int_{\gamma_{\mu}} f'(x', y') q_k^{(\mu)}(x', y') d\sigma,$$

where the functions $\{q_n^{(\mu)}(x', y')\}$ are defined by the relations

$$q_{2n}^{(\mu)}(x', y') = \bar{q}_{2n}(x'/\mu, y'/\mu) / \mu^{n+1}, \quad q_{2n+1}^{(\mu)}(x', y') = \bar{q}_{2n+1}(x'/\mu, y'/\mu) / \mu^{n+1}.$$

Hence for each value of μ , $0 < \mu < 1$, there is a set of functions $\{q_n^{(\mu)}(x', y')\}$ continuous and biorthogonal to the set $\{p'_n(x', y')\}$ on the circle γ_{μ} . And if we let $q''_n(x', y')$ be identified on each circle γ_{μ} with $q_n^{(\mu)}(x', y')$, we obtain a set of functions $\{q''_n(x', y')\}$ continuous and biorthogonal to the set $\{p'_n(x', y')\}$ on every circle γ_{μ} : $\rho = \mu$, $0 < \mu < 1$.

We have yet to show that the expansion (6) formed for a chosen value of μ converges continuously thruout the interior of γ . Let S be any closed subregion of the interior of γ . There is some number ν , $0 < \nu < 1$, such that both S and γ_{μ} are interior to the circle γ_{ν} : $\rho = \nu$. Then $f'(x', y')$ admits an expansion

$$(10) \quad f'(x', y') = \sum_{k=1}^{\infty} g_k p'_k(x', y'), \quad g_k = \int_{\gamma_{\nu}} f'(x', y') q''_k(x', y') d\sigma,$$

which converges uniformly on and within γ_{ν} and hence uniformly in S and on γ_{μ} . But (10) must be the same development as (6). For multiplication of these two series thru by $q''_k(x', y') d\sigma$ and integration term by term over γ_{μ} gives the equality of a_k and g_k .

This completes the proof of Theorem III for the case where $f'(x', y')$ is harmonic interior to the unit circle. The remaining two cases are handled in similar manner.

3. *Proof of Theorem I.* An arbitrary region of the type described in Theorem I can be mapped conformally onto the interior of the unit circle in the w -plane of Theorem III by means of analytic functions $w = \Phi(z)$ and $z = \Psi(w)$. The function $w = \Phi(z)$ may be defined on the boundary B of G so as to be continuous in the closed region $G + B$. We shall suppose that this has been done. Moreover each of the functions

$$w^n/\pi^{\frac{1}{2}} = [\Phi(z)]^n/\pi^{\frac{1}{2}} \quad (n = 1, 2, \dots),$$

can be approximated uniformly in the closed region $G + B$ as closely as desired by a polynomial $P_n(z)$,⁶ and when this approximation is taken sufficiently close $P_n(z)$ will have by Rouché's theorem⁷ precisely n zeroes in G . Moreover, these n zeroes may be made to coincide⁸ at the point of G which corresponds by the map to the origin in the (x', y') -plane. Let the polynomials $\{P_n(z)\}$ be so chosen and also so that

$$|P_n(z) - [\Phi(z)]^n/\pi^{\frac{1}{2}}| \leq \lambda^n \bar{\epsilon}_n, \quad z \text{ in } G + B,$$

where $\bar{\epsilon}_n$ are those of Theorem III. Consequently inequalities (5) will hold for $|w| < 1$, if we identify the functions $\{\pi_n(w)\}$ with the transforms of the polynomials $\{P_n(z)\}$. Accordingly the harmonic polynomials $\{p_n(x, y)\}$ are to be chosen as the set of real and imaginary parts of the polynomials $\{P_n(z)\}$ so that their transforms may be identified with the harmonic functions $\{p'_n(x', y')\}$ of Theorem III.

Theorem I now follows from Theorem III by virtue of the conformal map. For if $f(x, y)$ is harmonic in G , its transform $f'(x', y')$ is harmonic within the unit circle γ , and so $f'(x', y')$ may be expanded into a series in terms of the

⁶ Farrell, *American Journal of Mathematics*, vol. 54 (1932), pp. 571-578, Lemma in § 5.

⁷ See Bieberbach, *Lehrbuch der Funktionentheorie*, vol. 1, p. 185. The application of Rouché's theorem here is almost immediate. The function $\Omega_n(z) = (1/\pi^{\frac{1}{2}})[\Phi(z)]^n$ is certainly zero nowhere on the boundary of G . And if the approximating polynomial $P_n(z)$ be chosen so that $|P_n(z) - \Omega_n(z)| < 1/\pi^{\frac{1}{2}}$ for z on B , then we have $|P_n(z) - \Omega_n(z)| < |\Omega_n(z)|$ for z on B . It follows then by the theorem of Rouché that the two functions $\Omega_n(z)$ and $\Omega_n(z) + [P_n(z) - \Omega_n(z)]$ have the same number of zeroes in G , which means that $P_n(z)$ has just n zeroes in G .

⁸ See Walsh, *Transactions of the American Mathematical Society*, vol. 26 (1924), pp. 155-170, p. 164.

functions $\{p'_n(x', y')\}$ where the coefficients of the expansion are found by integration around an arbitrary circle γ_μ . The series converges continuously within γ . This means that $f(x, y)$ may be expanded into a series in terms of the harmonic polynomials $\{p_n(x, y)\}$, where the series converges continuously in G . In like manner the remaining two cases of Theorem I are readily verified. The functions $\{q_n(x, y)\}$ of Theorem I are to be defined by the relations

$$q_n(x, y) = q''_n(x', y') d\sigma/ds,$$

where the equation refers to points of C_μ and γ_μ which correspond under the conformal map, and where σ and s denote arc length along γ_μ and C_μ respectively.

UNION COLLEGE,
SCHENECTADY, NEW YORK.

CERTAIN PROBLEMS IN THE THEORY OF CLOSEST APPROXIMATION.¹

By PAUL G. HOEL.

Introduction. This paper is concerned with properties of minimizing sums for integrals containing the m -th power of the error of an approximation, primarily for values of $m < 1$. By a minimizing sum is meant a linear combination of a set of n suitably restricted, linearly independent functions $s_1(x), s_2(x), \dots, s_n(x)$ which minimizes the integral

$$(1) \quad \int_a^b w(x) |f(x) - \psi(x)|^m dx,$$

where $w(x)$ is a suitably restricted weight function, m is a given positive constant, $\psi(x) = c_1 s_1(x) + c_2 s_2(x) + \dots + c_n s_n(x)$ is an arbitrary linear combination of the s 's, and $f(x)$ is a given suitably restricted function not identically equal to a linear combination of the s 's, to which it is desired to approximate by means of the ψ 's.

Three properties of minimizing sums are considered here. The first two are extensions, with appropriate modifications in statement, of properties of minimizing polynomials which are well known for an exponent $m > 1$, while the discussion of the third is believed to be new for $m \geq 1$, as well as for $m < 1$. The paper is divided into three sections corresponding to these three properties.

Throughout this paper, except where explicitly stated otherwise, the function $f(x)$ and the properly independent functions $s_1(x), s_2(x), \dots, s_n(x)$ are assumed to be bounded and measurable on the interval (a, b) , and the weight function $w(x)$ to be summable and non-negative on (a, b) , but positive over a subset of positive measure. The functions $s_1(x), s_2(x), \dots, s_n(x)$ are said to be properly independent on (a, b) if every linear combination of them, in which the coefficients are not all zero, is different from zero on a subset of positive measure. Under these restrictions, it can be shown by the usual methods of proof² that there exists at least one linear combination of the s 's which minimizes (1). Let $\phi_m(x)$ denote such a minimizing sum. For $m < 1$ there is not in general a unique minimizing sum; therefore, it will be assumed that $\phi_m(x)$ represents any one of the minimizing sums, if more than one exists.

¹ Presented to the American Mathematical Society, June 23, 1933.

² See, for example, D. Jackson, "A generalized problem in weighted approximation," *Transactions of the American Mathematical Society*, vol. 26 (1924), pp. 133-154; p. 137.

1. *The orthogonality property.* The purpose of this section is to demonstrate that the orthogonality property of a minimizing sum holds, within certain restrictions, when the exponent $m \leq 1$. In the case of $m > 1$, D. Jackson³ has shown that this property is both necessary and sufficient for a continuous minimizing sum of a continuous function. Here the condition will be shown to be a necessary one for this more general class of minimizing sums and functions, but naturally under more severe restrictions when $m \leq 1$.

THEOREM 1. For $m < 1$ the expression $w(x) \frac{|f(x) - \phi_m(x)|^m}{f(x) - \phi_m(x)}$, provided that both it and the expression with $w(x)$ deleted are summable on (a, b) , is orthogonal over the interval (a, b) to every linear combination $\psi(x)$ of the s 's; that is,

$$\int_a^b w(x) \frac{|f(x) - \phi_m(x)|^m}{f(x) - \phi_m(x)} \psi(x) dx = 0.$$

Proof. Let $I(h) = \int_a^b w |r - hs|^m dx$, where $r \equiv f - \phi_m$, s is any one of the s 's, and h is a constant. Then $I(0)$ represents the minimized integral. Without restricting the problem, it may be assumed that $|s(x)| \leq 1$, for the assumption can be realized by defining a new set of s 's which are the proper constant multiples of the old set. Now form

$$(2) \quad \frac{I(h) - I(0)}{h} = \int_a^b w \frac{|r - hs|^m - |r|^m}{h} dx.$$

Let E_1 , E_2 , and E_3 be the subsets of (a, b) for which respectively $r > 0$, $r < 0$, and $r = 0$. Then the derivative of $I(h)$ at the point $h = 0$ will exist and will be given by

$$(3) \quad I'(0) = \lim_{h \rightarrow 0} \frac{I(h) - I(0)}{h} = \lim_{h \rightarrow 0} \int_{E_1 + E_2 + E_3} w \frac{|r - hs|^m - |r|^m}{h} dx,$$

if the limit exists for the integrals over E_1 , E_2 , and E_3 separately. Only the limit as h approaches zero through positive values will be discussed, but the method is identical for the limit from the left.

From the hypothesis of the theorem that $|r|^m/r$ is summable on (a, b) , it follows that the measure of the set E_3 is zero. Consequently, the integral over E_3 vanishes and its limit is zero.

In considering the integral over E_1 , let $h_1, h_2, \dots, h_i, \dots$ be an arbitrary set of positive numbers for which $\lim_{i \rightarrow \infty} h_i = 0$. Let $g_1, g_2, \dots, g_i, \dots$ denote

³ "On functions of closest approximation," *ibid.*, vol. 22 (1921), pp. 117-128; p. 126.

the corresponding values of the integrand in (3). This set of functions is a sequence which satisfies the following three conditions almost everywhere on E_1 .

(a) For every value of h_i , $g_i(x)$ is obviously summable on E_1 .

(b) $\lim_{i \rightarrow \infty} g_i(x) = g(x)$, if, by definition, $g(x) \equiv -w(x)ms(x)[r(x)]^{m-1}$.

For, if x_0 represents an arbitrary point of E_1 , it is possible, from the definition of the set E_1 , to select an h_i so small that $h_i < r(x_0)/2$. Then for $h \leq h_i$, since $|s(x)| \leq 1$, $|r(x_0) - hs(x_0)|^m = [r(x_0) - hs(x_0)]^m$ is a positive differentiable function of h ; consequently,

$$\lim_{h \rightarrow 0} \frac{|r(x_0) - hs(x_0)|^m - |r(x_0)|^m}{h} = \left\{ \frac{d}{dh} [r(x_0) - hs(x_0)]^m \right\}_{h=0} = -ms(x_0)[r(x_0)]^{m-1}.$$

This limit holds in particular as h approaches zero through the succession of values h_i ; therefore,

$$\lim_{i \rightarrow \infty} g_i(x_0) = -w(x_0)ms(x_0)[r(x_0)]^{m-1} = g(x_0).$$

(c) $|g_i(x)| \leq G(x)$, if, by definition, $G(x) \equiv 10w(x)[r(x)]^{m-1}$. For, when $r(x) > 2h_i$ and $h \leq h_i$, the law of the mean may be applied to the expression below, considered as a differentiable function of h , to give

$$(4) \quad |r - hs|^m - |r|^m = [r - hs]^m - [r]^m = -msh[r - \theta_1 hs]^{m-1}, \\ 0 < \theta_1 < 1.$$

Furthermore, $[r - \theta_1 h_i s]^{m-1} < [r/2]^{m-1} < 2[r]^{m-1}$. Consequently, because of (4)

$$|g_i(x)| = |wms[r - \theta_1 h_i s]^{m-1}| \leq 2mw|s|[r]^{m-1} \leq 2w[r]^{m-1}.$$

When $r(x) \leq 2h_i$,

$$|g_i(x)| = w \left| \frac{|r - h_i s|^m - |r|^m}{h_i} \right| \leq w \frac{(3h_i)^m + (2h_i)^m}{h_i} \leq 10w[r]^{m-1}.$$

Hence (c) holds for all x and i , where $G(x)$, by hypothesis, is summable on E_1 .

Now, by a well known theorem on Lebesgue integration, the fact that $\{g_i(x)\}$ satisfies the above three conditions justifies the limit

$$(5) \quad \lim_{i \rightarrow \infty} \int_{E_1} w \frac{|r - h_i s|^m - |r|^m}{h_i} dx = -m \int_{E_1} ws[r]^{m-1} dx.$$

Since $\{h_i\}$ is an arbitrary sequence with limit zero, (5) implies that

$$\lim_{h \rightarrow 0} \int_{E_1} w \frac{|r - hs|^m - |r|^m}{h} dx = -m \int_{E_1} ws[r]^{m-1} dx.$$

The methods used to obtain this limit may be applied to the integral in (3) over E_2 to obtain a similar result. Combination of these three limit results gives,

$$I'(0) = -m \int_{E_1} ws \frac{|r|^m}{r} dx - m \int_{E_2} ws \frac{|r|^m}{r} dx = -m \int_a^b ws \frac{|r|^m}{r} dx.$$

Now $I(h)$ is a continuous function of h which has a minimum at $h = 0$; and since it possesses a derivative at that point, the derivative must vanish there. Since s represents any one of the s 's, this means that

$$\int_a^b ws_i \frac{|r|^m}{r} dx = 0 \quad (i = 1, 2, \dots, n),$$

which is equivalent to the vanishing integral of the theorem.

For $m = 1$ the details of the proof are simpler while the theorem hypotheses merely require that the measure of E_s be zero.

If $w(x)$ is positive on (a, b) , in addition to its original restrictions, Theorem 1 will be found to hold without the summability restriction on $|r|^m/r$.

It should be noted in the case of polynomial approximation that when the orthogonality property holds, it gives at once the information that $r(x)$ must possess at least n sign-changes in (a, b) ; since otherwise $\psi(x)$ could be chosen as a polynomial of degree $\leq n - 1$ with the same sign as $|r|^m/r$ to contradict the theorem. With n replaced by $2n + 1$, the same statement can be made concerning trigonometric approximation.

2. *Zero multiplicity bounds of the error function $f(x) - \phi_m(x)$.* This section is concerned with polynomial approximation to a continuous function. When $f(x)$ is continuous and $m > 1$, it is known⁴ that the error function corresponding to a minimizing polynomial of degree $n - 1$ changes sign at least n times in the interval. When $f(x)$ is continuous and $m = 1$, it is known⁵ that the error function changes sign at least n times or else vanishes over a set of positive measure. When $f(x)$ is analytic and $m < 1$, it is known⁶ that the sum of the multiplicities of the zeros of the error function must be at

⁴ Loc. cit., "On functions of closest approximation."

⁵ D. Jackson, "Note on a class of polynomials of approximation," *Transactions of the American Mathematical Society*, vol. 22 (1921), pp. 320-326.

⁶ D. Jackson, "Note on the convergence of a sequence of approximating polynomials," *Bulletin of the American Mathematical Society*, vol. 37 (1931), pp. 69-72; p. 72.

least n . The problem here is to make as definite a statement as possible when $m < 1$ and $f(x)$ is assumed to be merely continuous.

In this section let $w(x)$ be bounded, in addition to previous restrictions. As before let $r(x) = f(x) - \phi_m(x)$ denote a minimizing error function, where now $\phi_m(x)$ represents a minimizing polynomial of degree $\leq n-1$. Since $r(x)$ need not have zeros of definite multiplicities, a similar concept will be introduced by means of the following definition.

If x_k is a zero of $r(x)$, q_k will be called its multiplicity bound provided that for every $q > q_k$ there exists a positive constant δ_q such that the inequality

$$(8) \quad |r(x)| \geq |x - x_k|^q$$

holds for all values of x for which $|x - x_k| < \delta_q$, while if $q < q_k$ no such δ_q can be found. It can readily be shown that there exists a unique q_k , finite or infinite, for every zero of a continuous function.

THEOREM 2. *For $0 < m < 1$ the number of zeros and $(1-m)$ times the sum of the zero multiplicity bounds of $r(x)$ must add up to at least n .*

Assume the theorem to be false. Let t represent the number of zeros and u the sum of the zero multiplicity bounds. Then $t + (1-m)u < n$. Let x_1, x_2, \dots, x_t denote the zeros and q_1, q_2, \dots, q_t the corresponding multiplicity bounds. It is possible then to construct a polynomial $p(x)$ of proper degree which has the same sign as $r(x)$ everywhere in (a, b) , with $|p(x)| \leq 1$ in (a, b) , and with zeros of multiplicity p_k satisfying the relation

$$(9) \quad q_k(1-m) - 1 < p_k \leq q_k(1-m) + 1.$$

More explicitly $p(x)$ is expressed by

$$p(x) = C(x - x_1)^{p_1}(x - x_2)^{p_2} \cdots (x - x_t)^{p_t},$$

where C is a constant. The range of two units in the inequality (9) makes it possible to choose p_k odd or even according as $r(x)$ does or does not change sign at $x = x_k$. From (9) it follows that

$$(10) \quad \sum_{k=1}^t p_k \leq \sum_{k=1}^t [q_k(1-m) + 1] \leq (1-m)u + t < n.$$

This inequality shows that $p(x)$ is of degree $\leq n-1$. Now form

$$(11) \quad \begin{aligned} J(h) &= \int_a^b w |r - hp|^m dx, \\ J(h) - J(0) &= \int_a^b w [|r - hp|^m - |r|^m] dx, \end{aligned}$$

where h is a positive constant to be specified later. Let E_1 , E_2 , and E_3 be the subsets of (a, b) for which respectively $|r| > 2h|p|$, $h|p|/2 \leq |r| \leq 2h|p|$, and $|r| < h|p|/2$. By construction, $p(x)$ has the same sign as $r(x)$; consequently integral (11) is non-positive over the sets E_1 and E_2 and non-negative over E_3 . Moreover, since $J(0)$ is the minimum value of (1), integral (11) must be non-negative for every h . Hence

$$\int_{E_3} w[|r - hp|^m - |r|^m] dx \geq \int_{E_1 + E_2} w[|r|^m - |r - hp|^m] dx.$$

A contradiction will be obtained by showing that there exists a value of h for which this inequality is reversed. It will suffice to prove that

$$(12) \quad \int_{E_1} w[|r|^m - |r - hp|^m] dx > \int_{E_3} w[|r - hp|^m - |r|^m] dx.$$

The mean value theorem may be applied to the integrand on the left to give

$$\begin{aligned} |r|^m - |r - hp|^m &= mh|p| [|r| - \theta_2 h|p|]^{m-1} \quad (0 < \theta_2 < 1), \\ &= mh|p| |r|^{m-1} [1 - \theta_2 h|p|/|r|]^{m-1} > mh|p| |r|^{m-1}. \end{aligned}$$

Hence,

$$\int_{E_1} w[|r|^m - |r - hp|^m] dx \geq mh \int_{E_1} w|p| |r|^{m-1} dx.$$

Consider the integral on the right of (12). Since $|r| < h|p|/2$ on E_3 ,

$$\int_{E_3} w[|r - hp|^m - |r|^m] dx \leq \int_{E_3} w|hp|^m dx.$$

By combining these two results, it is seen that (12) is true if

$$(13) \quad m \int_{E_1} w|p| |r|^{m-1} dx > \int_{E_3} w|p| |hp|^{m-1} dx.$$

Let h be given a sufficiently small positive value h_0 such that the set E_1 will contain a subset of positive measure on which w is positive. The function $|p| |r|^{m-1}$ has but a finite number of zeros; consequently the integral on the left of (13) is positive with h so chosen. Let its value be denoted by K/m . Since the measure of E_1 is non-decreasing as h decreases, this integral is non-decreasing as h decreases. That means, for $h \leq h_0$,

$$(14) \quad m \int_{E_1} w|p| |r|^{m-1} dx \geq K(h_0) \equiv K > 0.$$

Over the set E_3 , $|hp| > 2|r|$; hence

$$|p| |hp|^{m-1} < |p| |2r|^{m-1} < |p| |r|^{m-1},$$

and

$$(15) \quad \int_{E_3} w |p| |hp|^{m-1} dx \leq \int_{E_3} w |p| |r|^{m-1} dx.$$

The integrand on the right becomes infinite only at the zeros of $r(x)$. Therefore to prove that the integral exists, it will suffice to show that it converges at any such zero x_k in E_3 . The demonstration follows.

It is possible from (9) to find a positive number $\rho < m$ such that $p_k > q_k(1-m) - 1 + \rho$. If $\epsilon > 0$ is chosen so small that $\epsilon(1-m) < \rho$, then

$$(16) \quad p_k > (q_k + \epsilon)(1-m) - 1.$$

From definition (8) there exists a $\delta_\epsilon < 1$ such that $|r(x)| \geq |x - x_k|^{q_k + \epsilon}$ for $|x - x_k| < \delta_\epsilon$. Moreover, from the definition of $p(x)$, there must exist a positive constant R such that $|p(x)| \leq R|x - x_k|^{p_k}$ for $|x - x_k| < \delta_\epsilon$. Combination of these last inequalities gives

$$(17) \quad |p| |r|^{m-1} \leq R|x - x_k|^{(q_k + \epsilon)(m-1) + p_k} \text{ for } |x - x_k| < \delta_\epsilon.$$

From (16) it follows that the exponent in (17) is > -1 . Consequently the integral on the right of (15) converges at x_k , and hence throughout E_3 . In view of (14) and (15), it is evident that (13) is true if

$$K > \int_{E_3} w |p| |r|^{m-1} dx.$$

If now h is allowed to approach zero, the measure of E_3 approaches zero; and therefore this last integral approaches zero with h . But K is independent of h ; consequently an h can be selected so small that the inequality holds. This proves the theorem.

COROLLARY 1. *If the number of zeros is not more than nm , the sum of the zero multiplicity bounds must be at least n .*

THEOREM 3. *If every $q_k < 1/(1-m)$, the number of sign changes of $r(x)$ in (a, b) is at least n .*

Proof. Here $q_k(1-m) + 1 < 2$ and $q_k(1-m) - 1 < 0$; hence, from (9), $p_k = 1$, if $r(x)$ changes sign at x_k , and $p_k = 0$, if it does not. If the theorem is assumed to be false, (10) becomes $\sum p_k \leq n - 1$, and the proof follows from there on as in Theorem 2. Under this hypothesis, there can be only a finite number of roots since at a limit point of roots q_k is infinite.

THEOREM 4. *The conclusions of Theorems 2 and 3 and Corollary 1 hold if $w(x)$ is not bounded but is such that $w(x)|r(x)|^{m-1}$ is summable on (a, b) .*

Proof. The details of the proof are identical with those of the above theorems through (15), after which the conclusion follows readily from the added summability hypothesis.

3. *Continuity of coefficients.* The problem of determining the behavior of a minimizing sum as the exponent m becomes infinite has been investigated in several papers.⁷ The main result of these investigations has been to show that in the case of polynomial approximation to a continuous function the minimizing polynomial, for $m \rightarrow \infty$, approaches the Tchebycheff polynomial of best approximation as a limit. However, no attention seems to have been given to the corresponding problem of determining the behavior of a minimizing sum as the exponent m approaches a finite value. The purpose of this section is to investigate that problem to the extent of showing that under suitable restrictions the coefficients of a minimizing sum are continuous functions of the exponent m .

In addition to the original set of restrictions, let $w(x)$ be non-vanishing almost everywhere on (a, b) . It will be assumed that $|f(x)| \leq 1$, since if the condition is not fulfilled originally, the said coefficients merely need to be multiplied by a suitable constant factor. It will also be assumed that $f(x)$ and the s 's form a set of $n + 1$ properly independent functions; for, otherwise, the minimizing sum is identical with $f(x)$ almost everywhere on (a, b) , for every m , and the problem of dependence on m becomes trivial. First, a necessary lemma will be proved.

LEMMA. *The absolute values of the coefficients of $\phi_m(x)$ have an upper bound independent of m for $\alpha \leq m \leq \beta$, where α and $\beta > \alpha$ represent any two positive numbers.*

Proof. Let J represent the integral (1) and $\bar{J} = \int_a^b w(x) |\theta(x)|^m dx$, where $\theta(x)$ is an arbitrary linear combination of $f(x)$ and the s 's with its least upper bound $= 1$. Then, by the same proof as that given in a paper by D. Jackson,⁸ it can be shown that

$$(18) \quad |c_k| \leq B(J/A_m)^{1/m},$$

where B is a constant depending only on $f(x)$ and the s 's, and A_m is the minimum value of \bar{J} . When $J \leq A_m$, $|c_k| \leq B$; otherwise, $|c_k| \leq B(J/A_m)^{1/a}$; and so in all cases

⁷ G. Pólya, "Sur un algorithme . . .," *Comptes Rendus*, vol. 157 (1913), pp. 840-843; D. Jackson, *loc. cit.*, "On functions of closest approximation"; J. Shohat, "On the polynomial of the best approximation to a given continuous function," *Bulletin of the American Mathematical Society*, vol. 31 (1925), pp. 509-514.

⁸ *Loc. cit.*, "A generalized problem in weighted approximation."

$$(19) \quad |c_k| \leq B[1 + (J/A_m)^{1/\alpha}].$$

For a fixed θ , since $|\theta| \leq 1$, $|\theta|^m$ is non-increasing as m increases, for each x , and so $\int_a^b w |\theta|^m dx$ is a non-increasing function of m . Hence, for every θ , $\int_a^b w |\theta|^m dx \geq \int_a^b w |\theta|^\beta dx \geq A_\beta$. Therefore $A_m \geq A_\beta$.

When only minimizing sums are considered,

$$J = \int_a^b w |f - \phi_m|^m dx \leq \int_a^b w |f|^m dx \leq \int_a^b w dx,$$

since $|f(x)| \leq 1$ and zero may be regarded as a particular linear combination of the s 's.

When these inequalities are applied to (19), $|c_k|$ for $\phi_m(x)$ will be found to have an upper bound independent of m over the range (α, β) .

As a consequence of this lemma and the fact that $f(x)$ and the s 's are bounded, there must exist a $K > 1$ independent of m such that

$$(20) \quad |f(x) - \phi_m(x)| < K \quad \text{for } \alpha \leq m \leq \beta.$$

THEOREM 5. *The coefficients of $\phi_m(x)$ are continuous functions of the exponent m for every value of $m > 0$.*

Proof. When $m > 1$, $\phi_m(x)$ is uniquely determined for each value of m . When $m \leq 1$, in the absence of further restrictions, this is not necessarily the case. The property of continuity then requires special explanation, which will be given later. Introduce the notations:

$$\begin{aligned} I(m, c^m) &= \int_a^b w(x) |f(x) - \phi_m(x)|^m dx \equiv \int_a^b w |r|^m dx \\ I(\mu, c^\mu) &= \int_a^b w(x) |f(x) - \phi_\mu(x)|^\mu dx \equiv \int_a^b w |r|^\mu dx \\ I(\mu, c^\mu) &= \int_a^b w(x) |f(x) - \phi_\mu(x)|^\mu dx \equiv \int_a^b w |z|^\mu dx \\ I(m, c^\mu) &= \int_a^b w(x) |f(x) - \phi_\mu(x)|^m dx \equiv \int_a^b w |z|^m dx. \end{aligned}$$

Since $I(m, c^m)$ and $I(\mu, c^\mu)$ are minimized values for exponents m and μ respectively,

$$(21) \quad I(m, c^m) \leq I(m, c^\mu); \quad I(\mu, c^\mu) \leq I(\mu, c^m).$$

Consider the difference

$$|I(\mu, c^\mu) - I(m, c^\mu)| = \left| \int_a^b w [|z|^\mu - |z|^m] dx \right|.$$

This difference does not exceed a quantity which may be written as

$$(22) \quad \int_a^b w |z|^m ||z|^{\mu-m} - 1| dx, \quad \text{or} \quad \int_a^b w |z|^\mu ||z|^{m-\mu} - 1| dx,$$

according as $\mu > m$ or $\mu < m$. Both of these integrals have the limit zero as μ approaches m . This is shown as follows.

Let ϵ be any positive quantity such that $\epsilon_0 = (\epsilon/4W)^{2/a} < 1$, where $W = \int_a^b w(x) dx$. Let E_1 , E_2 , and E_3 be the subsets of (a, b) for which respectively $|z| \leq \epsilon_0$, $\epsilon_0 < |z| \leq 1$, and $|z| > 1$. Choose $|\mu - m| < a/2$ and consider for $\mu < m$ the last integral in (22) in the three parts corresponding to these three sets. For the first part

$$\begin{aligned} \int_{E_1} w |z|^\mu ||z|^{m-\mu} - 1| dx &\leq \int_{E_1} w |z|^\mu dx \\ &\leq \int_{E_1} w |z|^{a/2} dx \leq \epsilon_0^{a/2} \int_{E_1} w dx \leq \epsilon/4. \end{aligned}$$

For the second and third parts, because of (20),

$$\int_{E_2+E_3} w |z|^\mu ||z|^{m-\mu} - 1| dx \leq K^\beta \int_{E_2+E_3} w ||z|^{m-\mu} - 1| dx.$$

The two parts of the integral on the right satisfy the inequalities

$$\begin{aligned} \int_{E_2} w ||z|^{m-\mu} - 1| dx &= \int_{E_2} w [1 - |z|^{m-\mu}] dx \leq (1 - \epsilon_0^{|\mu-m|}) W. \\ \int_{E_3} w ||z|^{m-\mu} - 1| dx &= \int_{E_3} w [|z|^{m-\mu} - 1] dx \leq (K^{|\mu-m|} - 1) W. \end{aligned}$$

Identical results hold for the first integral in (22). Combination of these results gives

$$|I(\mu, c^\mu) - I(m, c^\mu)| \leq \epsilon/4 + WK^\beta [(1 - \epsilon_0^{|\mu-m|}) + (K^{|\mu-m|} - 1)].$$

Since ϵ_0 and K are independent of $|\mu - m|$, the second term on the right can be made $\leq \epsilon/4$ by a sufficiently small choice of $|\mu - m|$, say δ . Hence,

$$(23) \quad |I(\mu, c^\mu) - I(m, c^\mu)| \leq \epsilon/2 \quad \text{for } |\mu - m| < \delta.$$

Since m and μ may take on any positive values, they are interchangeable here; hence,

$$(24) \quad |I(m, c^m) - I(\mu, c^m)| \leq \epsilon/2 \quad \text{for } |\mu - m| < \delta.$$

When μ differs from m by less than δ , (21) and (24), and (23) and (21)

combine to give the inequalities $I(\mu, c^\mu) \leq I(\mu, c^m) \leq I(m, c^m) + \epsilon/2$ and $I(\mu, c^\mu) \geq I(m, c^\mu) - \epsilon/2 \geq I(m, c^m) - \epsilon/2$. These are equivalent to

$$(25) \quad |I(\mu, c^\mu) - I(m, c^m)| \leq \epsilon/2 \quad \text{for } |\mu - m| < \delta.$$

Finally, combining (23) and (25) gives: $|I(m, c^m) - I(m, c^\mu)| \leq \epsilon$ for $|\mu - m| < \delta$. But this says that

$$(26) \quad \lim_{\mu \rightarrow m} I(m, c^\mu) = I(m, c^m).$$

The continuity proof results in interpreting this limit as follows. Let μ_i ($i = 1, 2, \dots$) be an arbitrary set of values of μ such that $\lim_{i \rightarrow \infty} \mu_i = m$.

Then from (26)

$$(27) \quad \lim_{i \rightarrow \infty} I(m, c^{\mu_i}) = I(m, c^m).$$

When $m > 1$, there corresponds to each value of μ_i a unique set of coefficients c^{μ_i} , which may be thought of as the coördinates of a point in an n -dimensional space. From the lemma, these points are bounded and thus give rise to at least one limit point as i becomes infinite. Since $I(m, c^\mu)$ is a continuous function of c^μ , the limiting value of $I(m, c^{\mu_i})$, as i becomes infinite, is the value taken on for the set corresponding to a limit point. But uniqueness for $m > 1$, because of (27), requires that the set of coefficients corresponding to a limit point be identical with c^m . Hence, there can be but *one* limit point, and $\lim_{i \rightarrow \infty} c^{\mu_i} = c^m$. Since μ_i is an arbitrary set approaching m , this is sufficient to prove continuity because it implies that

$$(28) \quad \lim_{\mu \rightarrow m} c^\mu = c^m.$$

When $m \leq 1$, the functions $I(m, c^\mu)$ and c^μ are to be thought of as multiple valued functions of μ possessing as many values as there are minimizing sums for that exponent. The reader will find, upon carefully examining the details of the last paragraph, that the same type of argument applies to arrive at (28), the limit to exist in this multiple valued sense. Here the continuity referred to in the theorem means that for every $\epsilon > 0$ there exists a $\delta > 0$ such that for every μ for which $|\mu - m| < \delta$, every ϕ_μ will have coefficients differing from those of some one of the ϕ_m^\bullet by less than ϵ .

REGULAR CONVERGENCE AND MONOTONE TRANSFORMATIONS.

By G. T. WHYBURN.

In a recent paper¹ the author defined the concept of regular convergence for a sequence of sets relative to r -dimensional cycles. A convergent sequence of closed sets $[A_n]$ is said to converge *regularly relative to r -dimensional cycles*, or simply *r -regularly*, provided that for every $\epsilon > 0$ there exist positive numbers δ and N such that if $n > N$ then any r -dimensional complete (Vietoris) cycle in A_n of diameter $< \delta$ is ~ 0 in a subset of A_n of diameter $< \epsilon$. It was shown that in the case of every type of set there studied (i. e., simple arcs, simple closed curves, topological spheres and closed 2-cells), regular convergence relative to 0-cycles for a sequence $[A_n]$ yields as a limiting set A a set either of the same topological type as the members of the sequence or a set which can be produced by an upper semi-continuous decomposition² of such a member into continua. In other words, the limiting set is always of such a nature that it can be represented as the image under a monotone transformation³ of a member of the sequence. This suggests the possibility of an intimate connection existing between regular convergence and monotone transformations, and in the present paper a theorem giving the exact nature of this relationship will be established (see Theorem 2 below).

We shall assume that all sets considered are imbedded in a compact metric space. The distance between two points x and y is denoted by $\rho(x, y)$, the diameter of a set X by $\delta(X)$ and the ϵ -neighborhood of a set X by $V_\epsilon(X)$, i. e., $V_\epsilon(X)$ is the set of all points y such that for some $x \in X$, $\rho(x, y) < \epsilon$. We use the arrow \rightarrow to indicate convergence rather than to show bounding relationships. Thus $A_n \rightarrow A$ means that $\limsup A_n = \liminf A_n = A$. All our cycles are non-oriented. A complete (Vietoris) r -dimensional cycle will be denoted by γ^r . A closed set A is said to be locally γ^r -connected⁴ provided that

¹ See my paper "On sequences and limiting sets," *Fundamenta Mathematicae*, vol. 25 (1935).

² See R. L. Moore, "Foundations of point set theory," *American Mathematical Society Colloquium Publications*, 1932, Chapter V.

³ See C. B. Morrey, *American Journal of Mathematics*, vol. 57 (1935), pp. 17-50, particularly p. 26.

⁴ Compare with Alexandroff, *Annals of Mathematics*, ser. 2, vol. 30 (1928), p. 181; and Lefschetz, "Topology," *American Mathematical Society Colloquium Publications*,

for every $\epsilon > 0$ there exists a $\delta > 0$ such that any γ^r in A of diameter $< \delta$ is ~ 0 in a subset of A of diameter $< \epsilon$. For a systematic treatment of the combinatorial notions used the reader is referred to the papers of Vietoris and Alexandroff.⁵

THEOREM 1. *If the sequence of closed sets $[A_n]$ converges to the limiting set A regularly relative to s -cycles for every $s \leq r$, then A is locally γ^r -connected.*

Proof. By virtue of a previous theorem of the author's⁶ we have only to show that for $\epsilon > 0$ there exists a $d > 0$ such that, for any $\epsilon > 0$, positive numbers δ and N exist such that if $n > N$, then any r -dimensional δ -cycle in A_n of diameter $< d$ is $\sim_\epsilon 0$ in a subset of A_n of diameter $< \epsilon$. To this end let $\epsilon > 0$ be given. By virtue of the r -regular convergence there exist positive numbers d and N_r with $3d < \epsilon/3$ such that if $n > N_r$, then any γ^r in A_n of diameter $< 3d$ is ~ 0 in a subset of A_n of diameter $< \epsilon/3$.

Now let $\epsilon > 0$ be given and let $\delta_r = \min(\epsilon, d)$. By virtue of the $(r-1)$ -regular convergence, there exist positive numbers δ_{r-1} and N_{r-1} such that if $n > N_{r-1}$, then any γ^{r-1} in A_n of diameter $< 3\delta_{r-1}$ is ~ 0 in a subset of A_n of diameter $< \delta_r$. Likewise, by virtue of the $(r-2)$ -regular convergence, using δ_{r-1} we have positive numbers δ_{r-2} and N_{r-2} such that if $n > N_{r-2}$, then any γ^{r-2} in A_n of diameter $< 3\delta_{r-2}$ is ~ 0 in a subset of A_n of diameter $< \delta_{r-1}$. Similarly, using δ_{r-2} , the $(r-3)$ -regular convergence gives us numbers δ_{r-3} and N_{r-3} , and so on. Continuing in this way we reach numbers δ_0 and N_0 such that if $n > N_0$, then any γ^0 in A_n of diameter $< 3\delta_0$ is ~ 0 in a subset of A_n of diameter $< \delta_1$. We may suppose the δ 's chosen so that $\delta_r > \delta_{r-1} > \dots > \delta_0$.

Let us now take $\delta = \delta_0$ and $N = \sum_0^r N_i$. We shall show that this choice of δ and N satisfies the required conditions for the given ϵ relative to d and ϵ .

To this end let n be any integer $> N$ and let C^r be any r -dimensional δ -cycle in A_n of diameter $< d$. Let $\Delta = (x_0, x_1, \dots, x_r)$ be any r -simplex in C^r . Now since $\delta(\Delta) < \delta_0$, it follows by 0-regular convergence that each 1-dimensional side (x_i, x_j) of Δ bounds a 1-dimensional semi-cycle⁷ x_{ij} in A_n

1930, pp. 91-92. See also the author's paper in the *American Journal of Mathematics*, vol. 56 (1934), pp. 133-146.

⁵ See, for example, Vietoris, *Mathematische Annalen*, vol. 97 (1927), pp. 454-472, and *Fundamenta Mathematicae*, vol. 19 (1932), pp. 265-273; Alexandroff, *loc. cit.*, pp. 101-187, and *Mathematische Annalen*, vol. 106 (1932), pp. 161-238.

⁶ See my paper "On sequences and limiting sets," *loc. cit.*, Theorem (1.4).

⁷ If $s > 0$ and $Z^{s-1} = (z_1, z_2, \dots)$ is a complete cycle of dimension $s-1$, then by an s -dimensional semi-cycle bounded by Z^{s-1} is meant a sequence of s -dimensional complexes $C^s = (c_1, c_2, \dots)$ where c_i is a σ_i -complex bounded by z_i , where $\sigma_i \rightarrow 0$ and where, for any given $\epsilon > 0$, $c_i \sim_\epsilon c_{i+j}$ for i sufficiently large, in the sense that if K_i is

of diameter $< \delta_1$. Now for each 2-dimensional side (x_i, x_j, x_k) of Δ , $c_{ijk} = x_{ij} + x_{ik} + x_{jk}$ is a γ^1 in A_n of diameter $< 3\delta_1$. Hence by 1-regular convergence, since $N > N_1$, it follows that c_{ijk} bounds a 2-dimensional semi-cycle x_{ijk} in A_n of diameter $< \delta_2$. Similarly, for each 3-dimensional side (x_i, x_j, x_k, x_m) of Δ , $c_{ijkm} = x_{ijk} + x_{ijm} + x_{ikm} + x_{jkm}$ is a γ^2 in A_n of diameter $< 3\delta_2$. Thus c_{ijkm} bounds a 3-dimensional semi-cycle x_{ijkm} in A_n of diameter $< \delta_3$. Continuing in this manner we eventually obtain, for each $(r-1)$ -dimensional side $(x_{i_0}, x_{i_1}, \dots, x_{i_{r-1}})$ of Δ a γ^{r-2} , $c_{i_0 i_1 \dots i_{r-1}}$ in A_n of diameter $< 3\delta_{r-2}$. By virtue of the $(r-2)$ -regular convergence, since $N > N_{r-2}$, this γ^{r-2} bounds an $(r-1)$ -dimensional semi-cycle $x_{i_0 i_1 \dots i_{r-1}}$ in A_n of diameter $< \delta_{r-1}$. Then $c_{12 \dots r} = \Sigma x_{i_0 i_1 \dots i_{r-1}}$ is a γ^{r-1} in A_n of diameter $< 3\delta_{r-1}$. Accordingly it bounds an r -dimensional semi-cycle $x_{12 \dots r}$ in A_n of diameter $< \delta_r$. The sum of all such semi-cycles $x_{12 \dots r}$ for all r -simplexes Δ in C^r is a γ^r , say Z , in A_n of diameter $< d + 2\delta_r \leq 3d$. Furthermore, if we write $Z = (z_1, z_2, \dots)$ as a γ^r , it is readily seen that we have $C^r \underset{\delta_r}{\sim} z_k$ in a subset of A_n of diameter $< d + 2\delta_r \leq 3d < e/3$, for every k . Whence, $C^r \underset{\epsilon}{\sim} z_k$ for all k , since $\delta_r \leq \epsilon$.

Now since $\delta(Z) < 3d$ and since $n > N > N_r$, it follows by r -regular convergence that $Z \underset{\epsilon}{\sim} 0$ in a subset of A_n of diameter $< e/3$. Thus for k sufficiently large, we have $z_k \underset{\epsilon}{\sim} 0$ in a subset of A_n of diameter $< e/3$. Therefore we have $C^r \underset{\epsilon}{\sim} z_k \underset{\epsilon}{\sim} 0$ in a subset of A_n of diameter $< e$, as was to be proven.

COROLLARY. *Under the hypothesis of Theorem 1, A is locally γ^s -connected for every $s \leq r$.*

Definition. If A is compact and metric and the transformation $T(A) = B$ is continuous, then T is said to be r -monotone provided that for each $b \in B$ all the connectivity (Betti) numbers of $T^{-1}(b)$ of dimension $\leq r$ vanish. For the case $r = 0$ this reduces to the ordinary monotone transformation as defined by Morrey (*loc. cit.*).

THEOREM 2. *Let the sequence of r -monotone transformations $T_n(A) = B_n$ converge uniformly to the limit transformation $T(A) = B$. In order that the sequence $[B_n]$ converge to B s -regularly for every $s \leq r$ it is necessary and sufficient that T be r -monotone and that B be locally γ^s -connected for every $s \leq r$.*

Proof. To prove the conditions necessary, suppose that $B_n \rightarrow B$ s -regularly

any s -dimensional complex of sufficiently small norm bounded by $z_i + z_{i+j}$, the s -cycle $K_i + c_i + c_{i+j}$ bounds an $(s+1)$ -dimensional ϵ -complex.

for every $s \leq r$. Let s be any integer $\leq r$, let $b \in B$, let $K = T^{-1}(b)$ and let $\epsilon > 0$ be given. We shall show first that every γ^s in K is ~ 0 in $V_\epsilon(K)$.

Now there exists a neighborhood R of b such that $T^{-1}(B \cdot \bar{R}) \subset V_{\epsilon/2}(K)$. There exists an integer N_1 such that for every $n > N_1$, $T_n^{-1}(B_n \cdot \bar{R}) \subset V_\epsilon(K)$. Also, by virtue of the s -regular convergence, there exists an N_2 such that for every $n > N_2$, $T_n(K) \subset R$ and every γ^s in $T_n(K)$ is ~ 0 in $B_n \cdot R$. Now let C be any γ^s in K . Let n be a fixed integer $> N_1 + N_2$. Then $T_n(C)$ is a γ^s in $B_n \cdot \bar{R}$; and since $T_n(C) \subset T_n(K)$, we have $T_n(C) \sim 0$ in $B_n \cdot \bar{R}$. Now, taking $B_n \cdot \bar{R} = \mathfrak{R}$ and $T_n^{-1}(B_n \cdot \bar{R}) = K \subset V_\epsilon(K)$ and applying Vietoris' result (10),⁸ it follows that since $T_n(C) \sim 0$ in $B_n \cdot \bar{R}$, C must be ~ 0 in $T_n^{-1}(B_n \cdot \bar{R})$. Whence, $C \sim 0$ in $V_\epsilon(K)$.

Now from this it results at once that any such cycle C in K must be ~ 0 in K . For let $C = (z_1, z_2, \dots)$ and let $\epsilon > 0$ be given. By what was just shown it follows that for k sufficiently large, z_k bounds an $\epsilon/3$ -complex K_0^{s+1} in $V_{\epsilon/3}(K)$; and clearly this complex projects by an $\epsilon/3$ -projection⁹ into an ϵ -complex K^{s+1} in K bounded by z_k . Thus we have shown that any complete cycle C in K of dimension $\leq r$ is ~ 0 in K , and accordingly T is r -monotone.

Finally, that B is locally γ^s -connected for every $s \leq r$ is a consequence of Theorem 1. Hence the conditions are necessary.

To prove the sufficiency of the conditions, we suppose the contrary. It follows that for some $s \leq r$, some $\epsilon > 0$ and some infinite sequence of integers (m_n) we have carriers $C_n \subset B_{m_n}$ of complete s -dimensional cycles Z_n such that $C_n \rightarrow b \in B$ but such that for no n is $Z_n \sim 0$ in any subset of $B_{m_n} \cdot V_{2\epsilon}(b)$.

We note first that if $W_n(x) = T_n T^{-1}(x)$, for each $x \in B$, then $[W_n(x)]$ converges uniformly to the identity transformation $W(x) = x$ on B . Now for the given ϵ , let us determine $\delta < \epsilon$ as given by the local γ^s -connectivity of B . Then by the uniform convergence $W_n(x) \rightarrow W(x)$, it follows that there exists an N such that if $n > N$ then $W_n(x) \subset V_{\delta/2}(x)$ for every $x \in B$. This gives

$$(i) \quad W_n[B \cdot V_{\delta/2}(b)] \subset V_\delta(b) \quad \text{and} \quad W_n^{-1}[B_n \cdot V_{\delta/2}(b)] \subset V_\delta(b);$$

and since $\epsilon > \delta$, we have likewise

$$(ii) \quad W_n[B \cdot V_\epsilon(b)] \subset V_{2\epsilon}(b).$$

Now let us determine an integer k so that $m_k > N$ and $C_k \subset V_{\delta/2}(b)$. Then since $C_k \subset B_{m_k}$, (i) gives $W_{m_k}^{-1}(C_k) \subset V_\delta(b)$. Now by Vietoris' result (6)

⁸ See *Mathematische Annalen*, vol. 97 (1927), p. 469.

⁹ That is, for each vertex x_0 in K_0^{s+1} which is not already in K , we substitute a point x of K with $\rho(x, x_0) < \epsilon/3$.

[*loc. cit.*, using $\mathfrak{K} = C_k$ and $K = T_{m_k}^{-1}(C_k)$], it follows that $T_{m_k}^{-1}(C_k)$ is an s -dimensional complete cycle Z'_k such that $T_{m_k}(Z'_k) \sim Z_k$ in B . Since $T(Z'_k) \subset V_\delta(b)$, it follows that $T(Z'_k) \sim 0$ in $B \cdot V_\epsilon(b)$. The r -monotone, it follows by Vietoris' result (10) (*loc. cit.*) that $T^{-1}[B \cdot V_\epsilon(b)]$; and by (ii) and the continuity of T_{m_k} it follows that $T_{m_k}(Z'_k) \sim 0$ in $B_{m_k} \cdot V_{2\epsilon}(b)$. But since $Z_k \sim T_{m_k}(Z'_k)$ in $C_k \subset B_{m_k} \cdot V_\delta(b)$, this gives $Z_k \sim 0$ in $B_{m_k} \cdot V_{2\epsilon}(b)$, contrary to the assumption. Thus the supposition that the conditions are not sufficient leads to a contradiction, and the theorem is proven.

COROLLARY. *If the sequence of r -monotone transformations T_n converges uniformly to the limit transformation $T(A) = B$ and B is γ^s -connected for every $s \leq r$, then T is r -monotone.*

For, in the case of a sequence of the type $[B, B, B, \dots]$, it is seen that s -regular convergence of the sequence and local γ^s -connectedness of the set B are equivalent.

Note. Let $r = 0$ and let us suppose all the transformations T_n are homeomorphisms. Consider the following choices for A : (1) the point $(0, 1)$, (2) the circle $x^2 + y^2 = 1$, (3) the sphere $x^2 + y^2 + z^2 = 1$, (4) the closed circular region $x^2 + y^2 \leq 1$. Then the sequence $T_n(A)$ represents, respectively, a sequence of (1) simple arcs, (2) simple closed curves, (3) spheres and (4) closed 2-cells. Now Theorem 2 tells us that if $T_n(A)$ converges 0-regularly, then the transformation $T(A) = B$ is monotone (= r -monotone) and thus by known results,¹⁰ in these cases, B is respectively, a (a) point, (b) simple closed curve, (c) cactoid, (d) hemi-cactoid. Thus it is seen at first glance that we have obtained the principal end-results of the paper "On sequences and limiting sets" (*loc. cit.*) by direct application of the more general Theorem 2. However, this is far from the case, in order to obtain these conclusions about B in the cases (1)-(4), the "transformation" representations $T_n(A) = B_n$ of the sets A are entirely *superfluous*. It may very well be possible, given a convergent sequence $[B_n]$ of sets of types (1)-(4), to introduce transformation representations in such a way as to get uniform convergence until this can be demonstrated, the direct methods employed in the paper, even though restricted for the present to the lower dimensional cases, are much more effective in treating cases such as the ones here exemplified.

UNIVERSITY OF VIRGINIA.

¹⁰ See Moore, *loc. cit.*, and Morrey, *loc. cit.*

ON SOME ASYMPTOTIC RELATIONS FOR THE CHARACTERISTIC VALUES OF THE ELLIPTIC DIFFERENTIAL EQUATIONS.

By H. L. KRALL.

Let

$$L(u) = A(x, y)u_{xx} + 2B(x, y)u_{xy} + C(x, y)u_{yy} + D(x, y)u_x + E(x, y)u_y + F(x, y)u$$

be an elliptic differential expression defined in the region T which is bounded by a simple closed curve R . We assume that R has a continuously turning tangent and that the coefficients $A(x, y), B(x, y), \dots, F(x, y)$ have continuous partial derivatives of the first three orders in $T + R$.

Let $\lambda_1, \lambda_2, \dots (\lambda_i \leq \lambda_{i+1})$ be the C. V. (characteristic values) of the differential system

$$(A) \quad \begin{aligned} Lu &= \lambda v, & Mv &= \lambda u \\ [u]_{(x, y) \text{ on } R} &= [v]_{(x, y) \text{ on } R} = 0, \end{aligned}$$

where M is the operator adjoint to L and each C. V. is counted a number of times equal to its multiplicity.

In this paper we propose to show that ¹

$$\lambda_n \cong \frac{4\pi n}{\iint_T \frac{dx dy}{\sqrt{AC - B^2}}}.$$

This result was proved by Geppert ² [3] for the case $A = C = 1, B = 0$, and then from this result of Geppert's it was shown by Gheorghiu [4] that if $\sigma_1, \sigma_2, \dots (|\sigma_i| \leq |\sigma_{i+1}|)$ are the C. V. of the system

$$(B) \quad \begin{aligned} Lz + \sigma z &= 0 \\ [z]_R &= 0, \end{aligned}$$

the series

$$\sum_{n=1}^{\infty} \frac{1}{|\sigma_n|^{1+\epsilon}}$$

converges for every $\epsilon > 0$. We shall show that this result is also true for the general case.

¹ \cong is read "is asymptotically equal to."

² See the bibliography at the end of the paper. The number in square brackets indicates the particular article referred to.

Discussion of the system (A). By a solution of (A) we shall mean a pair of functions satisfying (A) and possessing continuous partial derivatives of the first two orders in $T + R$. These functions also satisfy the integral equations

$$\begin{aligned} (C) \quad u(x, y) &= \lambda \int_T \int \Gamma(x, y, \xi, \eta) v(\xi, \eta) d\xi d\eta \\ v(x, y) &= \lambda \int_T \int \Gamma(\xi, \eta, x, y) u(\xi, \eta) d\xi d\eta, \end{aligned}$$

and as well the equations

$$\begin{aligned} u(x, y) &= \lambda^2 \int_T \int \Gamma_1(x, y, \xi, \eta) u(\xi, \eta) d\xi d\eta; \\ \Gamma_1(x, y, \xi, \eta) &= \int_T \int \Gamma(x, y, r, s) \Gamma(\xi, \eta, r, s) dr ds, \\ v(x, y) &= \lambda^2 \int_T \int \Gamma_2(x, y, \xi, \eta) v(\xi, \eta) d\xi d\eta; \\ \Gamma_2(x, y, \xi, \eta) &= \int_T \int \Gamma(r, s, x, y) \Gamma(r, s, \xi, \eta) dr ds, \end{aligned}$$

where $\Gamma(x, y, \xi, \eta)$ is the Green's function (or extended Green's function) of the operator L with the boundary condition $[\Gamma(x, y, \xi, \eta)]_{(x, y) \text{ on } R} = 0$. The existence of such a function has been proven by Girard ([5] see especially page 322).

E. Schmidt [8] has shown that the C. V. of (C) are positive and that the corresponding solutions build a complete set. Hence any piecewise continuous function can be approximated in the mean by a Fourier series of the $\{u_n\}$ or $\{v_n\}$ where u_n, v_n are normalized, $\left(\int_T \int u_n^2 dx dy = 1\right)$, characteristic functions corresponding to λ_n .

THEOREM. *The C. V. of (A) are identical with the C. V. of*

$$\begin{aligned} (D) \quad MLu - \lambda^2 u &= 0 \\ [u]_R &= [Lu]_R = 0, \end{aligned}$$

and

$$\begin{aligned} (D') \quad LMv - \lambda^2 v &= 0 \\ [v]_R &= [Mv]_R = 0. \end{aligned}$$

Moreover the multiplicity of each C. V. is the same.

By a solution of (D) or (D') we shall mean a function satisfying the system and possessing continuous partial derivatives of the first three orders and piecewise continuous derivatives of the fourth order.

The solutions of (A) have continuous derivatives of the first four orders (Hopf [6], p. 210); accordingly, we may operate with M and L respectively, which shows that the solutions of (A) satisfy (D) and (D').

Now suppose that λ' is a C. V. of (D) but not of (A). Then

$$\begin{aligned}MLu' - \lambda'^2 u' &= 0 \\MLu_n - \lambda_n^2 u_n &= 0.\end{aligned}$$

Multiply the first by u_n , the second by u' , subtract and integrate, then

$$\begin{aligned}0 &= \int_T \int [u_n MLu' - u' MLu_n] dx dy - (\lambda_n^2 - \lambda'^2) \int_T \int u_n u' dx dy, \\&\int_T \int u_n u' dx dy = 0.\end{aligned}$$

Since the set $\{u_n\}$ is complete, it follows that $u' \equiv 0$ and λ' is not a C. V.

Suppose now that associated with the C. V. λ , there are n linear independent normalized $\left(\int_T \int u_n^2 dx dy = 1\right)$ solutions of (D) namely u_1, u_2, \dots, u_n , and t ($t \leq n$) linear independent solutions of (D'), v_1, v_2, \dots, v_t . We shall produce n linear independent solutions of (A) which will show that the multiplicity is the same throughout.

Let

$$\bar{u}_s = u_s, \quad \bar{v}_s = \sum_{j=1}^t a_{sj} v_j, \quad a_{sj} = (1/\lambda) \int_T \int v_j L(u_s) dx dy.$$

We wish to show that (\bar{u}_s, \bar{v}_s) is a solution of (A). Set

$$\begin{aligned}f_s &= L\bar{u}_s - \lambda \bar{v}_s = Lu_s - \lambda \sum_{j=1}^t a_{sj} v_j, \\g_s &= M\bar{v}_s - \lambda \bar{u}_s = \sum_{j=1}^t a_{sj} M(v_j) - \lambda u_s.\end{aligned}$$

Then

$$\begin{aligned}Mf_s &= -\lambda g_s, & LMf_s &= \lambda^2 f_s, \\Lg_s &= -\lambda f_s, & MLg_s &= \lambda^2 g_s.\end{aligned}$$

Accordingly $(-g_s, f_s)$ is a solution of (A) and f_s is a solution of (D') and we must have $f_s = \sum_{j=1}^t a_{sj} v_j$. Then

$$\begin{aligned}a_{sj} &= \int_T \int f_s v_j dx dy = \int_T \int v_j Lu_s dx dy - \lambda \sum_{k=1}^t a_{sk} \int_T \int v_j v_k dx dy \\&= \int_T \int v_j Lu_s dx dy - \lambda a_{sj} = 0.\end{aligned}$$

Since $f_s = 0$ we must have $g_s = 0$ and then (\bar{u}_s, \bar{v}_s) are solutions of (A). Moreover we must also have $n = t$, for otherwise there would exist non-zero constants $\{c_i\}$ such that $\sum_{i=1}^n c_i \bar{v}_i = 0$, and then

$$0 = \sum_{i=1}^n c_i M \bar{v}_i = \lambda \sum_{i=1}^n c_i \bar{u}_i,$$

which is a contradiction.

The maximum-minimum property of the C. V. In the following theorem (Courant-Hilbert [2, p. 326]), we use ϕ to denote a general function having continuous partial derivatives of the first three orders in $T + R$ and piecewise continuous partial derivatives in T .

THEOREM. *Let $\xi_1, \xi_2, \dots, \xi_{n-1}$ be arbitrary piecewise continuous functions in T , and let $g(\xi_1, \xi_2, \dots, \xi_{n-1})$ be the lower bound or minimum of*

$$K(\phi) = \int_T \int [L(\phi)]^2 d\sigma$$

for all admissible functions ϕ satisfying

$$\begin{aligned} \int_T \int \phi^2 d\sigma &= 1 \\ \int_T \int \phi \xi_i d\sigma &= 0 & (i = 1, 2, \dots, n-1) \\ [\phi]_R &= 0. \end{aligned}$$

Let λ_n be the n -th C. V. of (D). Then λ_n^2 is the maximum value which $g(\xi_1, \xi_2, \dots, \xi_{n-1})$ can assume, and this maximum-minimum value will be taken for $\xi_1 = u_1, \xi_2 = u_2, \dots, \xi_{n-1} = u_{n-1}, \phi = u_n$.

The Euler equation of this Calculus of Variations problem is $MLu - \lambda^2 u = 0$, and this equation must of course be satisfied by the minimizing function. We would not change the lower bound by demanding that $[L(\phi)]_R = 0$, because (Geppert [3]) we can approximate every ϕ by an admissible function ψ such that $[L(\psi)]_R = 0$ and $K(\phi)$ differs from $K(\psi)$ by an arbitrarily small amount.

To prove the theorem consider

$$w(P) = \sum_{i=1}^n c_i u_i(P), \quad (u_i(P) = u_i(x, y)).$$

The $\{c_i\}$ can be determined so that $w(P)$ is an admissible function. Then using Green's formula we have

$$K(w) = \int_T \int [L(w)]^2 d\sigma = \int_T \int wMLwd\sigma = \sum_{i=1}^n c_i^2 \lambda_i^2 \leq \lambda_n^2 \sum_{i=1}^n c_i^2 = \lambda_n^2.$$

Accordingly

$$g(\xi_1, \xi_2, \dots, \xi_{n-1}) \leq K(w) \leq \lambda_n^2.$$

Now take $\xi_1 = u_1, \xi_2 = u_2, \dots, \xi_{n-1} = u_{n-1}$. The function ϕ generates the series

$$\phi(P) \sim \sum_{i=1}^{\infty} \phi_i u_i(P), \quad (\phi_i = \int_T \phi u_i d\sigma).$$

Making use of Parseval's identity and the identities

$$\int_T \int fg d\sigma = \sum_{n=1}^{\infty} f_i g_i, \quad \int_T \int u_i M L \phi d\sigma = \int_T \int \phi M L u_i d\sigma = \lambda_i^2 \phi_i,$$

we have, since $\phi_1 = \phi_2 = \dots = \phi_{n-1} = 0$,

$$K(\phi) = \int_T \int \phi M L \phi d\sigma = \sum_{i=1}^{\infty} \lambda_i^2 \phi_i^2 \geq \lambda_n^2 \sum_{i=1}^{\infty} \phi_i^2 = \lambda_n^2.$$

But for $\phi = u_n$ we have $K(\phi) = \lambda_n^2$, which completes our proof.

Transformation of $K(\phi)$. Lichtenstein [7, p. 38] considered the expression

$$L_1(w) = A_1 w_{xx} + 2B_1 w_{xy} + C_1 w_{yy} + D_1 w_x + E_1 w_y + F_1 w$$

where $A_1 C_1 - B_1^2 = 1$. H showed that functions $X(x, y), Y(x, y)$ exist which are solutions of $(\partial/\partial x)(A_1 U_x + B_1 U_y) + (\partial/\partial y)(B_1 U_x + C_1 U_y) = 0$, and which transform the region in a one to one manner into the region T' , so that, setting $w(x, y) = W(X, Y)$

$$\begin{aligned} L_1(w) &= L'_1(W) = \delta(W_{XX} + W_{YY}) + D'W_X + E'W_Y + F'W \\ \delta &= A_1 X_x^2 + 2B_1 X_x X_y + C_1 X_y^2, \\ D' &= \{D_1 - (A_1)_x - (B_1)_y\}X_x + \{E_1 - (B_1)_x - (C_1)_y\}X_y \\ E' &= \{D_1 - (A_1)_x - (B_1)_y\}Y_x + \{E_1 - (B_1)_x - (C_1)_y\}Y_y \\ F' &= F_1. \end{aligned}$$

These functions $X(x, y), Y(x, y)$ must have continuous partial derivatives of the first four orders (Hopf [6], p. 210). The Jacobian of the transformation is δ , and

$$dx dy = (1/\delta) dX dY \quad (\delta > 0).$$

Setting $\alpha = (AC - B^2)^{1/2}$, $A_1 = A/\alpha, B_1 = B/\alpha, \dots, F_1 = F/\alpha$, we can write

$$K(\phi) = \int_T \int \alpha^2 [L_1(\phi)]^2 dx dy.$$

Thus the transformation gives us

$$\begin{aligned} K(\phi) = K(\Phi) &= \int_{T'} \int (\alpha^2/\delta) (\delta\Phi_{XX} + \delta\Phi_{YY} + D'\Phi_X + E'\Phi_Y + F'\Phi)^2 dXdY \\ &= \int_{T'} \int [\mathfrak{L}(\Phi)]^2 dXdY, \end{aligned}$$

where

$$\begin{aligned} \phi(x, y) &= \Phi(X, Y), & \mathfrak{L}(\Phi) &= p^{1/2}(\Phi_{XX} + \Phi_{YY}) + a\Phi_X + b\Phi_Y + f\Phi, \\ p^{1/2} &= \alpha\delta^{1/2}, & b &= \alpha E'/\delta^{1/2}, \\ a &= \alpha D'/\delta^{1/2}, & f &= \alpha F'/\delta^{1/2}. \end{aligned}$$

In the new variables the conditions on ϕ become

$$\begin{aligned} \int_T \int \phi^2 d\sigma &= \int_{T'} \int \rho \Phi^2 dXdY = 1 & (\rho = 1/\delta > 0), \\ \int_T \int \phi \xi_i d\sigma &= \int_{T'} \int \rho \Phi \xi_i dXdY = 0 & (i = 1, 2, \dots, n-1), \\ [\phi]_R &= [\Phi]_{R'} = [L(\phi)]_R = [\mathfrak{L}(\Phi)]_{R'} = 0. \end{aligned}$$

Accordingly

$$\lambda_n^2 = \text{Maximum-Minimum } K(\phi)$$

where the admissible functions Φ satisfy the above conditions.

In the same manner as before we can show that the $\{\lambda_n\}$ are the C. V. of the system

$$\begin{aligned} \mathfrak{M}\mathfrak{L}U - \lambda^2 \rho U &= 0 \\ (\text{E}) \quad [U]_{R'} &= [\mathfrak{L}(U)]_{R'} = 0 \\ \mathfrak{M} &= \text{operator adjoint to } \mathfrak{L}. \end{aligned}$$

Our proof of the maximum-minimum property required that the set $\{U_n\}$ of characteristic functions of (E) be complete. This completeness follows from the fact that the system (E) is equivalent to the integral equation

$$U(P) = \lambda^2 \int_{T'} \int \left[\int_{T'} G(P, N) G(Q, N) d\sigma_N \right] \rho(Q) U(Q) d\sigma_Q$$

where $G(P, Q)$ is the Green's function (ordinary or extended) of the operator L .

An asymptotic relation for the C. V. We start with a discussion of the system

$$(F) \quad \begin{aligned} \Delta(p \Delta W) - l^2 \rho W &= 0 \\ [W]_{R'} = [\Delta W]_{R'} &= 0 \end{aligned} \quad (\Delta = \partial^2/\partial X^2 + \partial^2/\partial Y^2)$$

where the p and ρ are defined above, and of course are greater than zero in T' . Then

$$l_n^2 = \text{Max-Min } H(\Phi) = \text{Max-Min} \int_{T'} \int p(\Delta \Phi)^2 dX dY$$

where the admissible functions Φ satisfy the same conditions as before.

Let $A(l)$ denote the number of C. V. less than or equal to l . Now if $p(P)$ is replaced by $p'(P)$ where $p' \geq p$ the minimums $h(\xi_1, \xi_2, \dots, \xi_{n-1})$ of $H(\Phi)$ cannot be decreased. Hence $l_n^2 = h(W_1, W_2, \dots, W_{n-1})$ cannot be decreased and $A(l)$ cannot be increased. If $\rho(P)$ is replaced by $\rho'(P)$ where $\rho' \leq \rho$ we introduce $\Phi' = k\Phi$ where k is such that

$$1 = \int_{T'} \int \rho' \Phi'^2 d\sigma = k^2 \int_{T'} \int \rho' \Phi^2 d\sigma \leq k^2 \int_{T'} \int \rho \Phi^2 d\sigma = k^2.$$

Consequently $k \geq 1$ and

$$H(\Phi') = k^2 H(\Phi) \geq H(\Phi).$$

Accordingly l_n cannot be decreased and $A(l)$ cannot be increased.

We now use the following lemma (Courant and Hilbert [2], p. 330).

LEMMA. Let T_1, T_2, \dots be non-overlapping sub-regions of T and let $A_{T_i}(l)$ designate the number of C. V. less than or equal to l for the region T_i , then

$$A(l) \geq A_{T_1}(l) + A_{T_2}(l) + \dots$$

We divide T' up into a network of squares T_i of side δ . In square T_i replace p by its maximum p_{M_i} and replace ρ by its minimum ρ_{m_i} , then $A_{T_i}(l)$ cannot be increased. Accordingly if $A'_{T_i}(l)$ refers to

$$\begin{aligned} p_{M_i} \Delta \Delta \Phi - l^2 \rho_{m_i} \Phi &= 0 \\ [\Phi]_R = [\Delta \Phi]_R &= 0, \end{aligned}$$

we must have

$$A_{T_i}(l) \geq A'_{T_i}(l).$$

But this last system is equivalent to the system

$$\begin{aligned} \Delta \Phi - l \sqrt{\frac{\rho_{m_i}}{p_{M_i}}} \Phi &= 0 \\ [\Phi]_R &= 0, \end{aligned}$$

and we have (Courant and Hilbert [2], p. 355)

$$A'_{T_i}(l) \cong \frac{\delta^2}{4\pi} \sqrt{\frac{\rho_{m_i}}{p_{M_i}}} l.$$

Hence

$$\begin{aligned} A(l) &\geq A'_{T_1}(l) + A'_{T_2}(l) + \cdots \\ &\cong \frac{l}{4\pi} \sum_i \delta^2 \sqrt{\frac{\rho_{m_i}}{p_{M_i}}} = \frac{l}{4\pi} \left\{ \int_{T'} \int \sqrt{\frac{\rho}{p}} d\sigma + \epsilon \right\} \\ A(l) &\cong \frac{l}{4\pi} \int_{T'} \int \sqrt{\frac{\rho}{p}} d\sigma; \quad l_n \cong \frac{4\pi n}{\int_{T'} \int \sqrt{\frac{\rho}{p}} d\sigma} \end{aligned}$$

To obtain an asymptotic lower bound for l_n set

$$S = \sqrt{\rho p}, \quad D(\Phi) = \int_{T'} \int S(\Phi x^2 + \Phi y^2) d\sigma$$

then

$$\begin{aligned} \sqrt{H(\Phi)} &= \sqrt{\int_{T'} \int p(\Delta\Phi)^2 d\sigma \int_{T'} \int \rho\Phi^2 d\sigma} \geq \left| \int_{T'} \int \Phi(S_x\Phi_x + S_y\Phi_y) d\sigma \right| \\ &= \left| D(\Phi) \right| - \left| \int_{T'} \int \Phi(S_x\Phi_x + S_y\Phi_y) d\sigma \right| \\ \sqrt{H(\Phi)} &\geq |D(\Phi)| - \left| \int_{T'} \int \Phi(S_x\Phi_x + S_y\Phi_y) d\sigma \right| \\ \sqrt{H(\Phi)} &\geq D(\Phi) - c D(\Phi)^{1/2} \quad (c = \text{a positive constant}) \end{aligned}$$

Let τ_n be the maximum-minimum value of $D(\theta)$ functions θ are continuous and have piecewise continuous satisfy

$$\begin{aligned} \int_{T'} \int \rho\theta^2 d\sigma &= 1 \\ \int_{T'} \int \rho\theta\xi_i d\sigma &= 0 \quad (i = 1, 2, \dots, n) \\ [\theta]_R &= 0. \end{aligned}$$

This $D(\theta)$ can be approximated by $D(\Phi)$ where Φ is a function of the former class. Hence τ_n is the n -th C. V. of the system

$$\begin{aligned} (S\theta_x)_x + (S\theta_y)_y + \tau\rho\theta &= 0 \\ [\theta]_R &= 0. \end{aligned}$$

We have the relation (Courant and Hilbert [2], p. 355)

$$\tau_n \cong \frac{4\pi n}{\int_{T'} \int (\rho/S) d\sigma}.$$

Now

$$\sqrt{h(\xi_1, \xi_2, \dots, \xi_{n-1})} = \text{Lower Bound } \sqrt{H(\Phi)} \geq \text{Lower Bound } [D(\Phi) - cD(\Phi)^{1/2}].$$

If we draw the graph of the curve $y = x - cx^{1/2}$ we see that the lower bound on the right is either less than zero or is taken on when $D(\Phi)$ is least. Hence if n is so large that $\tau_n > c^2$, we can find a set of $\{\xi_i\}$ so that lower bound of $D(\Phi)$ is greater than c^2 and for this set, the lower bound will be taken on when $D(\Phi)$ is least. Thus for $s \geq n$ we have

$$\begin{aligned} l_s &= \text{Max-Min } \sqrt{H(\Phi)} \\ &\geq \text{Minimum } \sqrt{H(\Phi)} \text{ for } \xi_i = \theta_i \quad (i = 1, 2, \dots, s-1) \\ &\geq \text{Minimum } [D(\Phi) - cD(\Phi)^{1/2}] \text{ for } \xi_i = \theta_i \quad (i = 1, 2, \dots, s-1) \\ &\geq D(\theta_s) - cD(\theta_s)^{1/2} = \tau_s - c\tau_s^{1/2} \\ l_s &\cong \frac{4\pi s}{\int_{T'} \int \sqrt{\frac{\rho}{p}} d\sigma}. \end{aligned}$$

We have the result

$$l_n \cong \frac{4\pi n}{\int_{T'} \int \sqrt{\frac{\rho}{p}} d\sigma}.$$

Now we return to the system (E) and look for the extremal of

$$K(\Phi) = \int_{T'} \int [\mathfrak{L}(\Phi)]^2 d\sigma.$$

Let c represent a constant depending on a, b, f, p and ρ , then

$$\begin{aligned} \left| \int_{T'} \int 2ap^{1/2}\Phi_X(\Delta\Phi) d\sigma \right| &\leq c \left\{ \int_{T'} \int \Phi_X^2 d\sigma \int_{T'} \int (\Delta\Phi)^2 d\sigma \right\}^{1/2} \\ &\leq c D(\Phi)^{1/2} H(\Phi)^{1/2} \\ \left| \int_{T'} \int 2fp^{1/2}\Phi(\Delta\Phi) d\sigma \right| &\leq c H(\Phi)^{1/2} \\ \left| \int_{T'} \int 2ab\Phi_X\Phi_Y d\sigma \right| &\leq c \int_{T'} \int (\Phi_X^2 + \Phi_Y^2) d\sigma \leq c D(\Phi). \end{aligned}$$

Making similar estimates for the other terms in K we obtain

$$K(\Phi) = H(\Phi) + O(D^{1/2} H^{1/2}) + O(H^{1/2}) + C$$

But

$$\sqrt{H} \geq D - c D^{1/2} = (D^{1/2} - c/2)^2 - c^2/4$$

$$D^{1/2} = O(H^{1/4})$$

and

$$K(\Phi) = H(\Phi) + O(H^{3/4}).$$

The terms in the expression $O(H^{3/4})$ are algebraic, derivative dK/dH will be positive for H sufficiently large. N exists such that for $H > N$, the minimum of $K(\Phi)$ $H(\Phi)$ takes on its minimum.

Using $k(\xi_1, \xi_2, \dots, \xi_{n-1})$ to denote the minimums of l bearing that

$$\lambda_n^2 = k(U_1, U_2, \dots, U_{n-1})$$

$$l_n^2 = h(W_1, W_2, \dots, W_{n-1}),$$

we have for n sufficiently large

$$l_n^2 - O(l_n^{3/2}) \leq k(W_1, W_2, \dots, W_{n-1}) \leq k(U_1, U_2, \dots, U_{n-1})$$

$$\lambda_n^2 \leq h(U_1, U_2, \dots, U_{n-1}) + O(h(U_1, U_2, \dots, U_{n-1})^{1/2})$$

$$\leq l_n^2 + O(l_n^{3/2}),$$

wherefore

$$\lambda_n \cong l_n \cong \frac{4\pi n}{\int \int_T \sqrt{\frac{\rho}{p}} dX dY}.$$

But $\rho = 1/\delta$, $p = \alpha^2 \delta$, $\alpha = \sqrt{AC - B^2}$, $dxdy = (1/\delta)dX$ forming to the variables x, y we have

$$\lambda_n \cong \frac{4\pi n}{\int \int_T \frac{dxdy}{\sqrt{AC - B^2}}}.$$

Consideration of the system (B).

$$(B) \quad \begin{aligned} Lz + \sigma z &= 0 \\ [z]_R &= 0. \end{aligned}$$

The solutions $z_n(x, y)$ satisfy the integral equation

$$z_n(x, y) = \sigma_n \int \int_T \Gamma(x, y, \xi, \eta) z_n(\xi, \eta) d\xi d\eta.$$

Hence the $\{\sigma_n\}$ are the zeros of the Fredholm determinant relative to the kernel $\Gamma(x, y, \xi, \eta)$. The C. V. λ_n are the zeros of Fredholm determinant relative to the kernels $\Gamma_1(x, y, \xi, \eta)$, $\Gamma_2(x, y, \xi, \eta)$ associated with $\Gamma(x, y, \xi, \eta)$.

Gheorghiu [4, p. 29] has proved that if the series $\sum_{n=1}^{\infty} (1/\lambda_n^\alpha)$ converges, the series $\sum_{n=1}^{\infty} (1/|\sigma_n|^\alpha)$ must converge, where of course the terms containing $\sigma_i = 0$ are omitted from the series. Setting $\alpha = 1 + \epsilon$ we have the result,

$$\sum_{n=1}^{\infty} (1/|\sigma_n|^{1+\epsilon})$$

converges for every $\epsilon > 0$.

PENNSYLVANIA STATE COLLEGE.

BIBLIOGRAPHY.

1. R. Courant, "Über die Schwingungen eingespannter Platten," *Mathematische Zeitschrift*, vol. 15, pp. 195-200.
2. R. Courant und D. Hilbert, *Methoden der Mathematischen Physik*, Erster Band, Berlin, 1924.
3. H. Geppert, "Über die Eigenwertprobleme bei nichtselbstadjungierten elliptischen Differentialgleichungen zweiter Ordnung," *Mathematische Annalen*, vol. 95 (1925), pp. 519-543.
4. D. A. Gheorghiu, *Sur l'équation de Fredholm*, Dissertation, Paris, 1928.
5. G. Girard, "Sur le problème de Dirichlet généralisé," *Annales de l'École Normale* (1929), pp. 131-245.
6. E. Hopf, "Über den funktionalen, insbesondere den analytischen Charakter der Lösungen elliptischer Differentialgleichungen zweiter Ordnung," *Mathematische Zeitschrift*, vol. 34 (1931), pp. 194-233.
7. L. Lichtenstein, "Randwertaufgaben der Theorie der linearen partiellen Differentialgleichungen zweiter Ordnung vom elliptischen Typus," *Journal für die reine und angewandte Mathematik*, vol. 142 (1913).
8. E. Schmidt, "Zur Theorie der linearen und nichtlinearen Integralgleichungen," *Mathematische Annalen*, vol. 63 (1907), pp. 433-476.

GEGENBAUER FUNCTIONS WITH A NEGATIVE INDEX.

By MAURICE DE DUFFAHEL.

Gegenbauer polynomials, which play a rather important part in several problems (for instance in the potential theory and harmonic analysis), are defined by the expansion

$$(1 - 2\alpha x + \alpha^2)^{-\nu} = \sum_{n=0}^{\infty} \alpha^n C_n^{\nu}(x)$$

where ν is a positive number. $C_n^{\nu}(x)$ is a Gegenbauer polynomial of order n , with the superior index ν . It can be expressed (omitting a constant factor) by means of a Gauss hypergeometric series as

$$F(-n/2, n/2 + \nu; 1/2; x^2)$$

if n is even, with a similar formula when n is odd.

Let us now suppose that ν is a *negative* number. Writing $\nu = -\mu$, we see that the hypergeometric function,

$$F(-n/2, n/2 - \mu; 1/2; x^2),$$

is definite, even if μ is an integer; but, in this case, the expansion

$$(1 - 2\alpha x + \alpha^2)^{\mu}$$

becomes improper, giving rise to a limited number of terms. The polynomial corresponding to the hypergeometric function, and which we shall name $C_n^{-\nu}(x)$, cannot therefore be attached to the said expansion. Now the question is, which generating function shall we have for this new polynomial?

The answer is, $C_n^{-\nu}(x)$ or $C_n^{\mu}(x)$ comes from the generating function

$$(1 - 2\alpha x + \alpha^2)^{\mu} \log(1 - 2\alpha x + \alpha^2).$$

In order to establish this result, let us begin with $\mu = 1$. Starting from

$$(1 - 2\alpha x + \alpha^2) \log(1 - 2\alpha x + \alpha^2) = \sum \alpha^n P_n(x),$$

we find readily that P_n satisfies the recurrence-formulae

$$\begin{aligned} nP_n &= xP'_n - P'_{n-1} \\ P'_{n+1} - xP'_n &= (n-2)P_n; \end{aligned}$$

hence the differential equation

$$(1-x^2)P''_n + xP'_n + n(n-2)P_n = 0.$$

But the differential equation for Gegenbauer polynomial C_n^ν is

$$(1-x^2)y'' - (2\nu+1)xy' + n(n+\nu-1)y = 0$$

and we obtain the differential equation for P_n by taking in the latter $\nu = -1$. So, omitting a constant factor, we infer immediately that

$$P_n = C_n^{-1}(x).$$

Coming now to the general case, if we have

$$(1-2\alpha x + \alpha^2)^\mu \log(1-2\alpha x + \alpha^2) = \sum_n \alpha^n P_n^\mu(x),$$

we can write

$$(1-2\alpha x + \alpha^2)^{\mu+1} \log(1-2\alpha x + \alpha^2) = \sum \alpha^n (1-2\alpha x + \alpha^2) \alpha^n P_n^\mu(x);$$

hence

$$P_n^{\mu+1} = P_n^\mu - 2\alpha P_{n-1}^\mu + P_{n-2}^\mu.$$

This being exactly one of Gegenbauer's recurrence formulae, we see that

$$P_n^2 = C_n^{-2}(x),$$

and so on.

$$* \quad * \quad * \quad * \quad *$$

This result can be used to find solutions of the partial differential equation with two variables

$$\Delta \cdots \Delta \Delta U = 0$$

or

$$\Delta^p U = 0,$$

where

$$\Delta U = \partial^2 U / \partial x^2 + \partial^2 U / \partial y^2.$$

For, if we write

$$u = x + iy \quad v = x - iy,$$

the equation becomes

$$\partial^{2p} U / \partial u^p \partial v^p = 0$$

of which the general solution is

$$U(x, y) = V(u, v) = u^{p-1}F(v) + v^{p-1}G(u),$$

F and G being arbitrary functions.

Let us take

$$F(v) = v^{p-1} \log v, \quad G(u) = u^{p-1} \log u,$$

so

$$U = w^{p-1} v^{p-1} \log uv,$$

or

$$U(x, y) = (x^2 + y^2)^{p-1} \log (x^2 + y^2).$$

This is a solution of $\Delta^p U = 0$, and it is clear that

$$\partial^m U / \partial x^m$$

is also a solution.

So also is the function

$$U_m(x, y) = \frac{(-1)^m}{m!} \frac{\partial^m}{\partial \alpha^m} \{[(x - \alpha)^2 + y^2]^{p-1} \log [(x - \alpha)^2 + y^2]\}.$$

But we can write, using Taylor's formula,

$$[(x - \alpha)^2 + y^2]^{p-1} \log [(x - \alpha)^2 + y^2] = \sum_m \alpha^m U_m(x, y).$$

Now suppose that we are on the circle

$$x^2 + y^2 = 1.$$

This formula becomes

$$(1 - 2\alpha x + \alpha^2)^{p-1} \log (1 - 2\alpha x + \alpha^2) = \sum \alpha^m U_m$$

so that U_m becomes identical to a Gegenbauer function with a negative index, here

$$C_m^{1-p}(x).$$

Gegenbauer functions with a negative index are therefore p -harmonic functions, in the xy -plane, and on the circle $x^2 + y^2 = 1$.

This fact is analogous to the fact of Legendre polynomials being harmonic functions, for space, on the sphere $x^2 + y^2 + \xi^2 = 1$, and Hermite two-variable polynomials being harmonic functions, for hyperspace, on the hypersphere $x^2 + y^2 + \xi^2 + z^2 = 1$.

STAMBOUL, TURQUIE D'EUROPE.

NOTE ON POTENTIAL THEORY IN n -SPACE.

By J. J. L. HINRICHSSEN.

In this note the elegant elementary methods developed by Schmidt¹ and applied by him to Newtonian potentials are extended to obtain the properties near the acting masses of the integrals defining the potentials of simple surface, double surface, and volume distributions in n -space.

The integrals by means of which we shall define surface and volume potentials in n -space ($n > 2$) are the following: for a simple distribution of density $f(x_1, \dots, x_{n-1})$ on a surface α :

$$(1) \quad V(\xi_1, \dots, \xi_n) = \int_{\alpha} f(1/r^{n-2}) d\sigma;$$

for a distribution of density $\rho(x_1, \dots, x_n)$ throughout a volume G :

$$(2) \quad A(\xi_1, \dots, \xi_n) = \int_G \rho(1/r^{n-2}) d\tau;$$

and for a double surface distribution of moment $f(x_1, \dots, x_{n-1})$ on α :

$$(3) \quad W(\xi_1, \dots, \xi_n) = \int_{\alpha} f \frac{\partial}{\partial \nu} (1/r^{n-2}) d\sigma.$$

Here $d\sigma$ and $d\tau$ represent the differential elements of surface and volume respectively, while the simple integral signs are used to denote the $(n-1)$ - and n -fold multiple integrals. Also

$$r^2 = \sum_{i=1}^n (\xi_i - x_i)^2,$$

and $\partial/\partial \nu$ represents the derivative in the direction of the normal to α . Exterior to the acting masses, these integrals define analytic functions of ξ_1, \dots, ξ_n which satisfy Laplace's equation in n -dimensions:

$$\nabla^2 u = \sum_{i=1}^n (\partial^2 u / \partial \xi_i^2) = 0.$$

¹ E. Schmidt, "Bemerkung zur Potentialtheorie," pp. 364-383 of *Mathematische Abhandlungen H. A. Schwarz gewidmet*, Berlin, 1914. Methods related to those of Schmidt are used by M. M. Sullivan in a paper entitled "On the derivatives of Newtonian and logarithmic potentials near the acting masses," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 137-171.

We shall investigate the behavior of these integrals as the point (ξ_1, \dots, ξ_n) approaches the distribution.

Let us assume that the surface containing the distribution or bounding the volume containing the distribution is finite and integrable. We shall suppose that the surface point P in whose neighborhood the potential is to be investigated is a simple point and that the surface has a continuous normal at P as well as in a certain neighborhood of P . That is to say, we assume the possibility of cutting from the surface a piece α which by proper choice of rectangular coördinates may be represented by:

$$(4) \quad x_n = F(x_1, \dots, x_{n-1}),$$

where $F, \partial F / \partial x_i$ are continuous for $|x_i| \leq a, a > 0$ and $i = 1, 2, \dots, (n-1)$. Also α shall contain P but not on its boundary, that is for $P; |x_i| < a$. Finally the remaining part of the surface after cutting out α cannot contain P either as a boundary or interior point. Since the potential at P of the remaining part of the surface is analytic, we need only investigate the potential of the distribution on α for the neighborhood of P .

We shall assume Green's theorem in n -dimensions:

$$\int_G \sum_{i=1}^n \frac{\partial X_i}{\partial x_i} d\tau = - \int_\gamma \sum_{i=1}^n X_i \cos(\nu, x_i) d\sigma, \quad (\text{inner normal}).$$

Then in a manner similar to that used for the Newtonian case, we can show if $\phi(x_1, \dots, x_n)$ is of class C''^2 throughout the region G and its boundary γ :

$$(5) \quad \int_G \sum_{i=1}^n \frac{\partial \phi}{\partial x_i} \frac{\partial}{\partial x_i} \left(\frac{1}{r^{n-2}} \right) d\tau = - \int_\gamma \phi \frac{\partial}{\partial \nu} \left(\frac{1}{r^{n-2}} \right) d\sigma + \begin{cases} (n-2) \mathcal{A} \phi(\xi_1, \dots, \xi_n) \\ \left(\frac{n-2}{2} \right) \mathcal{A} \phi(\xi_1, \dots, \xi_n), \\ 0 \end{cases}$$

and if ϕ is of class C'' ,

$$(6) \quad \int_G \frac{1}{r^{n-2}} \nabla^2 \phi d\tau = - \int_\gamma \frac{1}{r^{n-2}} \frac{\partial \phi}{\partial \nu} d\sigma + \int_\gamma \phi \frac{\partial}{\partial \nu} \left(\frac{1}{r^{n-2}} \right) d\sigma - \begin{cases} (n-2) \mathcal{A} \phi(\xi_1, \dots, \xi_n) \\ \left(\frac{n-2}{2} \right) \mathcal{A} \phi(\xi_1, \dots, \xi_n). \\ 0 \end{cases}$$

The first, second or third value in the braces is to be chosen according to

² By a function of class $C^{(i)}$ we mean a function which is continuous together with its partial derivatives up to the i -th order inclusive.

whether P is a point interior to G , a boundary point, or an exterior point respectively. Here

$$(7) \quad \mathcal{A} = \begin{cases} \frac{2(2\pi)^{(n-1)/2}}{1 \cdot 3 \cdot 5 \cdots (n-1)} & \text{if } n \text{ is odd,} \\ \frac{(2\pi)^{n/2}}{2 \cdot 4 \cdot 6 \cdots (n-2)} & \text{if } n \text{ is even.} \end{cases}$$

The lemmas corresponding to those given by Schmidt may be established for n -space by the same methods as for 3-space. It will be found that regardless of the position of the point P ,

$$\int_T (1/r^{n-1}) d\tau \leq K_1 v^{1/n}$$

$$\int_T (1/r^{n-2}) d\tau \leq K_2 v^{2/n}$$

$$\int_S (1/r^{n-2}) d\sigma \leq K_3 J^{1/(n-1)},$$

where K_1, K_2, K_3 are constants dependent on n , v is the n -dimensional volume of the region T , and J denotes the $(n-1)$ -dimensional volume of the projection of S on the $\xi_1, \xi_2, \dots, \xi_{n-1}$ -space. By means of these lemmas, it is not difficult to show that *with a bounded density, the potential of a simple surface distribution (1) is everywhere a continuous function of the position of the point P even when the point lies on the surface or on its boundary.* Likewise the same procedure may be followed to show that *with a bounded density, the potential of a finite volume distribution (2) and its first derivative is everywhere continuous,—also at a point of discontinuity of the density and on the boundary of the volume,—and the first derivative is obtainable from the integral defining the potential by differentiating under the integral sign.*

Consider now the piece of surface α defined by (4) with a double spread of moment $f(x_1, \dots, x_{n-1})$. Let us suppose that the functions f and F are of class C' . Choose h so large that the entire piece of surface α lies between the two planes $x_n = \pm h$ without having points in common therewith. The space bounded by $x_n = \pm h$, $x_i = \pm a$ ($i = 1, 2, \dots, n-1$), is then divided into two parts T_1 and T_2 by α . Let the positive normal on α point towards T_1 and γ_1 denote that part of the boundary of T_1 which is made up of planes bounding the n -dimensional parallelepiped.

We shall apply Green's theorem (5) to the region T_1 and the function $f(x_1, \dots, x_{n-1})$ which is defined inclusive of the boundaries of T_1 . Upon introducing W defined by (3), we obtain

$$(8) \quad W = - \int_{\gamma_1} f \frac{\partial}{\partial \nu} \left(\frac{1}{r^{n-2}} \right) d\sigma + \int_{T_1} \sum_{i=1}^{n-1} \frac{\partial f}{\partial x_i} \frac{\partial}{\partial \xi_i} \left(\frac{1}{r^{n-2}} \right) d\tau \\ + \begin{cases} (n-2) \mathcal{H} f(\xi_1, \dots, \xi_{n-1}) \\ \frac{n-2}{2} \mathcal{H} f(\xi_1, \dots, \xi_{n-1}). \\ 0 \end{cases}$$

On applying our earlier results to the integrals appearing in (8), we find that the potential of a double distribution (3) has continuous limit values on both sides of the surface which from the side of the negative to the side of the positive normal has a break of $(n-2) \mathcal{H} f$. On the surface itself, the defining integral has the value of the arithmetic mean of these two limit values.

Let us now suppose the function F to be of class C'' and f to be of class C' . Let Γ be the $p = (n-2)$ -dimensional manifold bounding α . Then Stokes' theorem in n -space³ may be written:

$$\int_{\Gamma} X_{s_1 \dots s_p} d(x_{s_1} \dots x_{s_p}) = \int_{\alpha} X_{s_1 \dots s_{p+1}} d(x_{s_1} \dots x_{s_{p+1}}),$$

where

$$X_{s_1 \dots s_{p+1}} = \frac{\partial X_{s_1 \dots s_p}}{\partial x_{s_{p+1}}} - \frac{\partial x_{s_{p+1}}}{\partial x_{s_1}} \frac{\partial X_{s_2 \dots s_p}}{\partial x_{s_1}} - \frac{\partial X_{s_1 s_{p+1} s_3 \dots s_p}}{\partial x_{s_2}} - \dots - \frac{\partial X_{s_1 \dots s_{p-1} s_{p+1}}}{\partial x_{s_p}}.$$

In the left member s_1, \dots, s_p are umbral and $s_1 < s_2 < \dots < s_p$, while in the right member s_1, \dots, s_{p+1} are umbral and $s_1 < s_2 < \dots < s_{p+1}$. Here s_1, s_2, \dots, s_p are any p numbers chosen from $1, 2, \dots, n$ and

$$X_{s_1 \dots s_i s_j \dots s_p} = -X_{s_1 \dots s_j s_i \dots s_p}.$$

For our purpose it is convenient to define

$$X_{s_1 s_2 \dots s_p} = \begin{cases} 0 & \text{if } s_1, \dots, s_p \text{ contains } j, \\ (-1)^{\bar{s}_p + j} f \cos(\nu, x_{s_p}^-) (1/r^p) & \text{if } s_1, \dots, s_p \text{ does not contain} \\ & j \text{ and } \bar{s}_p > j, \\ (-1)^{\bar{s}_p + j - 1} f \cos(\nu, x_{s_p}^-) (1/r^p) & \text{if } s_1, \dots, s_p \text{ does not contain} \\ & j \text{ and } \bar{s}_p < j. \end{cases}$$

Here $j, s_1, \dots, s_p, \bar{s}_p$ contains $1, 2, \dots, n$, and (ν, x_{s_p}) represents the angle between the normal to α and the x_{s_p} -axis. Upon substituting, we find it necessary to consider two cases, 1) n odd and 2) n even. Here as well as later we shall denote a function which is analytic provided that the point P does not lie on the boundary Γ of the piece of surface α by R .⁴

³ We shall adopt the notation of F. D. Murnaghan, *Vector Analysis and the Theory of Relativity* (1922), p. 33. Also see H. Poincaré, "Sur les résidues des intégrales doubles," *Acta Mathematica*, vol. 9 (1887), p. 321.

⁴ Even though two functions are not identical, if they have the above-mentioned property, they will both be denoted by R .

For case 1), we find

$$(9) \quad \frac{\partial V}{\partial \xi_j} = \int_a f_j \frac{1}{r^{n-2}} d\sigma - \int_a f \cos(\nu, x_j) \frac{\partial}{\partial \nu} \left(\frac{1}{r^{n-2}} \right) d\sigma - R,$$

where

$$f_j = \sum_{i=1}^n \left\{ \cos(\nu, x_i) \frac{\partial}{\partial x_j} [f \cos(\nu, x_i)] - \cos(\nu, x_j) \frac{\partial}{\partial x_i} [f \cos(\nu, x_i)] \right\}.$$

For the second case it is convenient to re-define the function f by setting

$$\bar{f} = f \sum_{i=1}^n (-1)^i \cos^2(\nu, x_i),$$

and then to remove the bar from the new function. We so obtain:

$$(9') \quad \frac{\partial V}{\partial \xi_j} = \int_a \bar{f}_j \frac{1}{r^{n-2}} d\sigma - \int_a \bar{f}_j \cos(\nu, x_j) \frac{\partial}{\partial \nu} \left(\frac{1}{r^{n-2}} \right) d\sigma - R,$$

where

$$\begin{aligned} \bar{f}_j = \sum_{i=1}^n \left\{ (-1)^i \cos(\nu, x_i) \frac{\partial}{\partial x_j} \left[\frac{f \cos(\nu, x_i)}{\sum_{i=1}^n (-1)^i \cos^2(\nu, x_i)} \right] \right. \\ \left. + (-1)^{j+1} \cos(\nu, x_j) \frac{\partial}{\partial x_i} \left[\frac{f \cos(\nu, x_i)}{\sum_{i=1}^n (-1)^i \cos^2(\nu, x_i)} \right] \right\} \end{aligned}$$

and

$$\bar{f}_j = \frac{(-1)^j f}{\sum_{i=1}^n (-1)^i \cos^2(\nu, x_i)}.$$

Applying our previous results to (9) and (9'), we find that the derivatives $\partial V / \partial \xi_j$ ($j = 1, 2, \dots, n$), have continuous limiting values on both sides of the surface α . The formulas for the break in limit values take on a simple form when the x_n -axis is chosen parallel to the normal through the point. Then we find $\partial V / \partial \xi_j$ ($j = 1, 2, \dots, n-1$), to be continuous while $\partial V / \partial \xi_n$ has a break of $-(n-2) \mathcal{A} f$ in going from the side of the negative to the side of the positive normal.

Now let the functions F and f be of class C'' . If we apply Green's theorem (6) to the region T_1 and the function $f(x_1, \dots, x_{n-1})$ defined in T_1 inclusive of its boundaries, we obtain

$$(10) \quad W = \int_{T_1} \frac{1}{r^{n-2}} \sum_{i=1}^{n-1} \frac{\partial^2 f}{\partial x_i^2} d\tau + \int_a \frac{1}{r^{n-2}} \sum_{i=1}^{n-1} \frac{\partial f}{\partial x_i} \cos(\nu, x_i) d\sigma + R \\ + \begin{cases} (n-2) \mathcal{A} f(\xi_1, \dots, \xi_{n-1}) \\ \frac{n-2}{2} \mathcal{A} f(\xi_1, \dots, \xi_{n-1}). \\ 0 \end{cases}$$

Upon analyzing the integrals in (10), we conclude that the derivatives $\partial W/\partial \xi_j$ ($j = 1, 2, \dots, n$), have continuous limit values on both sides of α . The formulas for the break in limit values take on a simple form when the x_n -axis is chosen parallel to the normal through the point. Then we obtain: $(n-2) \mathcal{H}(\partial f/\partial x_j)$ ($j = 1, 2, \dots, n-1$), in going from the side of the negative to the side of the positive normal. The derivative in the direction of the normal is continuous.

Let G be any region in n -space satisfying the conditions of Green's theorem. Let the boundary γ consist of γ_1 and α defined by (4) and let the density $\rho(x_1, \dots, x_n)$ be of class C' interior to and on the boundary of G . Then if A is defined by (2), it may be shown that

$$(11) \quad \frac{\partial A}{\partial \xi_j} = \int_G \frac{\partial \rho}{\partial \xi_j} \frac{1}{r^{n-2}} d\tau + \int_\alpha \rho \cos(\nu, x_j) \frac{1}{r^{n-2}} d\sigma + R.$$

We are now prepared to state the three general theorems.

THEOREM 1. *If F is of class $C^{(k)}$, ρ is of class $C^{(k-1)}$, then A together with its partial derivatives up to the k -th order inclusive are continuous both inside and outside of G and have continuous limit values on both sides of the bounding surface α .*

THEOREM 2. *If F is of class $C^{(k+1)}$, f is of class $C^{(k)}$, then V together with its partial derivatives up to the k -th order inclusive have continuous limit values on both sides of the surface α .*

THEOREM 3. *If F is of class $C^{(k+1)}$, f is of class $C^{(k+1)}$, then W together with its partial derivatives up to the k -th order inclusive have continuous limit values on both sides of the surface α .*

We have already outlined the proofs of these three theorems for $k=1$. By means of mathematical induction and the identities (11), (9), (9') and (10) we can show that these theorems are true for all values of k .

The formulas (11), (9) and (10) also yield the results for the successive breaks. For example, from (11) we may read off the breaks in the $(n+1)$ -th-order derivatives of the potential of a volume distribution on the boundary surface in terms of the breaks in the n -th-order derivatives of the potential of a volume distribution and the breaks in the n -th-order derivatives of the potential of a simple surface distribution. Equations (9) and (9') give the breaks on the surface in the $(n+1)$ -th-order derivatives of the potential of a simple surface distribution in terms of the breaks in the n -th-order derivatives

of the potential of a simple surface distribution and the breaks in the n -th-order derivatives of the potential of a double distribution. Finally from the breaks in the $(n+1)$ -th-order derivatives of the potentials due to a volume distribution and the breaks in the $(n+1)$ -th-order derivatives of the potential due to a simple surface distribution, we can find the break on the surface of the $(n+1)$ -th-order derivatives of the potential due to a double surface distribution.

Thus for example, if the x_n -axis is chosen parallel to the normal to the surface at P , we obtain the following breaks in the second-order derivatives $\partial^2 A / \partial \xi_j \partial \xi_k$:

$$\begin{aligned} & 0 && \text{if } k \neq n, \\ - (n-2) \mathcal{A} \rho \cos(v, x_j) && \text{if } k = n. \end{aligned}$$

If we define $\alpha_{jk} = (\partial/\partial x_k)(\cos(v, x_j))$, then $\alpha_{jn} = 0$ and the breaks in $\partial^2 V / \partial \xi_j \partial \xi_k$ are found to be:

$$\begin{aligned} & - (n-2) \mathcal{A} f \alpha_{jk} && \text{if } j, k \neq n \text{ and } n \text{ is odd,} \\ (-1)^{(j+1)} (n-2) \mathcal{A} f \alpha_{jk} && \text{if } j, k \neq n \text{ and } n \text{ is even,} \\ - (n-2) \mathcal{A} (\partial f / \partial x_j) && \text{if } j \neq k = n, \\ (n-2) \mathcal{A} f \sum_{i=1}^{n-1} \alpha_{ii} && \text{if } j = k = n. \end{aligned}$$

Likewise the breaks in $\partial^2 W / \partial \xi_j \partial \xi_k$ may be obtained:

$$\begin{aligned} & (n-2) \mathcal{A} (\partial^2 f / \partial x_j \partial x_k) && \text{if } j, k \neq n, \\ - (n-2) \mathcal{A} \sum_{i=1}^{n-1} (\partial f / \partial x_i) \alpha_{ij} && \text{if } j \neq k = n, \\ - (n-2) \mathcal{A} \sum_{i=1}^{n-1} (\partial^2 f / \partial x_i^2) && \text{if } j = k = n. \end{aligned}$$

Finally from the breaks in $\partial^2 A / \partial \xi_j^2$, we find that for an interior point of the region G ,

$$\nabla^2 A = - (n-2) \mathcal{A} \rho$$

which is Poisson's equation in n -dimensions. It is to be noted that for $n=3$ these results reduce to those given by Schmidt.

THE GENERALIZED DOUBLE PENDULUM.

By G. BAILEY PRICE.

Introduction. The dynamical system formed by suspending a simple pendulum from the bob of another simple pendulum and allowing the two to move in a vertical plane under the attraction of gravity is called a double pendulum [3¹; no results of this paper are used in the present treatment of the problem]. The following obvious generalization will be called a generalized double pendulum: a heavy particle of mass m_1 moves on a fixed, simple closed curve C_1 in a vertical plane and carries without rotation and in the same vertical plane a second simple closed curve C_2 , on which a second heavy particle of mass m_2 moves. The meanings of the terms n -pendulum and generalized n -pendulum are clear from these definitions also.

The purpose of this note is (1) to use the generalized n -pendulum to illustrate certain results in the general theory of reversible dynamical systems, and (2) to establish in the large a result for the generalized double pendulum which so far in the general theory has been established only in the small (see section 3 below).

1. *The equations of motion.* Take axes in the plane of C_1 and C_2 with the ξ_2 -axis directed downward and the ξ_1 -axis to the right. The curves C_1 , C_2 are defined parametrically by

$$(1) \quad C_1 : \xi_i = x_i(u), \quad C_2 : \xi_i = y_i(v), \quad (i = 1, 2),$$

where $x_i(u)$, $y_i(v)$ are analytic and periodic with periods ω_u , ω_v . Furthermore, u and v are arc lengths on C_1 and C_2 :

$$(2) \quad (x_u x_u) \equiv 1, \quad (y_v y_v) \equiv 1,$$

where the notation is explained by

$$(3) \quad (w_u z_v) \equiv \frac{\partial w_1}{\partial u} \frac{\partial z_1}{\partial v} + \frac{\partial w_2}{\partial u} \frac{\partial z_2}{\partial v}.$$

The coördinates of m_1 and m_2 are respectively

$$\xi_i = x_i(u), \quad \xi_i = x_i(u) + y_i(v), \quad (i = 1, 2).$$

¹Numbers in brackets refer to articles cited in the bibliography at the end of this note.

Then T and U are

$$(4) \quad \begin{aligned} T &= [(m_1 + m_2)/2][u'^2 + 2m(x_u y_v)u'v' + mv'^2], \\ U &= (m_1 + m_2)g[x_2(u) + my_2(v)], \end{aligned}$$

where $m = m_2/(m_1 + m_2)$. Then the equations of motion in the Lagrangian form give

$$(5) \quad \begin{aligned} u'' &= \frac{m(x_u y_v)(x_{uu} y_v)u'^2 - m(x_u y_{vv})v'^2 + g[x_u^{(2)} - m(x_u y_v)y_v^{(2)}]}{1 - m(x_u y_v)^2}, \\ v'' &= \frac{-(x_{uu} y_v)u'^2 + m(x_u y_{vv})(x_u y_v)v'^2 + g[y_v^{(2)} - (x_u y_v)x_u^{(2)}]}{1 - m(x_u y_v)^2}. \end{aligned}$$

Now $(x_u y_v)$ is the scalar product of two unit vectors by (2) and (3); then since $0 < m < 1$, we see that the denominators in (5) are positive. Then the equations of motion for the generalized double pendulum are analytic for all values of (u, v) .

2. *Application of general theory.* The generalized double pendulum is a reversible dynamical system with two degrees of freedom; we shall show now how certain results in the general theory of such systems can be applied to it.

There are two characteristic surfaces associated with the problem. The first of these is given by [5, equation (7)]. The system can be interpreted as the motion on this surface of a particle of unit mass under the action of forces derived from U in (4). The equations of this surface in (x_1, \dots, x_4) space are

$$x_i = m_1^{1/2}x_i(u), \quad x_{i+2} = m_2^{1/2}[x_i(u) + y_i(v)],$$

where $(i = 1, 2)$. Since $x_i(u)$, $y_i(v)$ are periodic, the first characteristic surface is homeomorphic to a torus. For the ordinary double pendulum it can be shown that this surface is a surface of revolution.

The second characteristic surface enters through the Principle of Least Action. It is defined by

$$(6) \quad ds^2 = (U + h)T dt^2,$$

and the orbits for a given value of h can be interpreted as geodesics on this surface [2, pp. 36-39]. Now for h sufficiently large, $(U + h)$ is positive for all values of (u, v) ; then since (6) is invariant under the group of transformations

$$\bar{u} = u + n_1\omega_u, \quad \bar{v} = v + n_2\omega_v,$$

where n_1, n_2 are any integers, we may consider that the second characteristic

surface is closed and homeomorphic to a torus. Then if the energy of the generalized double pendulum is large enough so that motion is possible over the entire first characteristic surface, the minimum and minimax methods [1, Part II] and their extensions [4] establish immediately the existence of an infinite number of closed periodic orbits.

Now the generalized double pendulum has a certain number of positions of equilibrium corresponding to a point on C_1 with one on C_2 at which the tangents to these curves are horizontal. There are at least four positions of equilibrium since each curve has at least one maximum and one minimum.

Assume now that

$$(7) \quad (x_u^{(2)})_0 = 0, \quad (x_{uu}^{(2)})_0 < 0, \quad (y_v^{(2)})_0 = 0, \quad (y_{vv}^{(2)})_0 < 0,$$

where the subscript zeros indicate that the derivatives are evaluated for $u = v = 0$. Then C_1 and C_2 have maxima for $u = v = 0$, and $(0, 0)$ is a position of stable equilibrium. We shall assume also that the parameters (u, v) are chosen in such a way that they increase as C_1, C_2 are described in the positive sense, where the rotation which carries the positive ξ_1 -axis into the positive ξ_2 -axis is positive. Then from (2) and (7) we have $(x_u y_v)_0 = 1$; and the expansions of T, U about $(0, 0; 0, 0)$ are

$$(8) \quad \begin{aligned} T &= [(m_1 + m_2)/2][u'^2 + 2mu'v' + mv'^2] + \dots, \\ U &= [(m_1 + m_2)g/2][(x^{(2)})_0 u^2 + m(y^{(2)})_0 v^2] + \dots \end{aligned}$$

Then it is possible to introduce principal coördinates [6, chap. VII] (q_1, q_2) so that T, U become

$$(9) \quad \begin{aligned} T &= (q_1'^2 + q_2'^2)/2 + \dots, \\ U &= (\lambda_1^2 q_1^2 + \lambda_2^2 q_2^2)/2 + \dots, \end{aligned}$$

where

$$(10) \quad \left. \begin{matrix} \lambda_1^2 \\ \lambda_2^2 \end{matrix} \right\} = g \frac{[(x_{uu}^{(2)})_0 + (y_{vv}^{(2)})_0] \pm [[(x_{uu}^{(2)})_0 + (y_{vv}^{(2)})_0]^2 - 4(1-m)(x_{uu}^{(2)})_0(y_{vv}^{(2)})_0]^{\frac{1}{2}}}{2(1-m)}.$$

Since (7) holds and $0 < m < 1$, we see that neither λ_1 nor λ_2 is zero, that both are pure imaginary, and that in general

$$(11) \quad \lambda_1 \neq \lambda_2.$$

It is now possible to apply the results of [5]. We introduce the parameter μ as in [5, section 3]. Since λ_1, λ_2 are pure imaginary, the limiting integrable system $\mu = 0$ has two fundamental periodic orbits O_1^0, O_2^0 . It follows from

(11) and [5, section 5, Corollary 1] that at least one of these orbits can be continued analytically for $\mu > 0$. Also the surface of section can be continued analytically, and the results of [5, section 10] can be applied. As long as the surface of section SS exists and the region of motion R is homeomorphic to a circular disc, there are at least two periodic orbits which join two points of the oval of zero velocity Z . Poincaré's Last Geometric Theorem can be applied at least if both O_1^0 and O_2^0 can be continued analytically; the rotation numbers α_B and α_P are functions of μ and the curves C_1 , C_2 , and we may conclude that at least for $\mu > 0$ but small there exists an infinite number of closed periodic orbits.

Suppose now that the curves C_1 , C_2 have the symmetry specified by the hypotheses

$$(12) \quad \begin{aligned} x_1(-u) &= -x_1(u), & y_1(-v) &= -y_1(v), \\ x_2(-u) &= x_2(u), & y_2(-v) &= y_2(v). \end{aligned}$$

Then from (4) we find that the system is symmetric in the origin on the characteristic surface. Furthermore, this symmetry is retained when principal coördinates are introduced; hence, the results of [5, section 11] can be applied. The ordinary double pendulum has the symmetry (12).

There are also positions of equilibrium at which $x_{uu}^{(2)}$ and $y_{vv}^{(2)}$ have opposite signs. Then (10) shows that one of the multipliers λ_1^2 , λ_2^2 is positive and the other negative; hence, the limiting system $\mu = 0$ has one periodic orbit, which can be continued analytically for $\mu > 0$. If $x_{uu}^{(2)}$ and $y_{vv}^{(2)}$ are both positive, then λ_1^2 , λ_2^2 are positive, and the limiting system has no periodic orbits.

The theorems of [5, section 5] yield entirely similar results for the generalized n -pendulum.

3. *On the existence of a periodic orbit joining two points of the oval of zero velocity.* The following question naturally arises: can the periodic orbits O_1 , O_2 [5, section 10] be continued indefinitely, or do they exist in the small only? We shall not attempt to answer this question, but we shall establish the existence in the large of at least one periodic orbit joining two points of the oval of zero velocity Z for the symmetric generalized double pendulum.

First we introduce the notation

$$(w_u z_v)^* = \begin{vmatrix} \frac{dw_1}{du} & \frac{dz_1}{dv} \\ \frac{dw_2}{du} & \frac{dz_2}{dv} \end{vmatrix}.$$

Then from the first relation in (2) we find that

$$[y_v^{(2)} - (x_u y_v) x_u^{(2)}] = x_u^{(1)} (x_u y_v)^*.$$

Then the second equation in (5) becomes

$$(13) \quad v'' = \frac{(x_{uu} y_v) u'^2 + m(x_u y_{vv}) (x_u y_v) v'^2 + g x_u^{(1)} (x_u y_v)^*}{1 - m(x_u y_v)^2}.$$

We now assume that our generalized double pendulum is symmetric in the origin, i. e., we assume that (12) holds. Then there is a position of equilibrium at $(0, 0)$, and we assume that it is stable as specified by (7). Let G be defined as the connected region of the (u, v) plane about the origin in which the following hypotheses are satisfied:

$$(14) \quad (x_{uu} x_{uu}) > 0, \quad (y_{vv} y_{vv}) > 0;$$

$$(15) \quad x_u^{(1)} < 0, \quad y_v^{(1)} < 0;$$

$$(16) \quad \begin{aligned} x_u^{(2)} &= 0 \quad \text{only for } u = 0, \\ y_v^{(2)} &= 0 \quad \text{only for } v = 0; \end{aligned}$$

$$(17) \quad (x_u y_v) > 0.$$

All of these conditions are satisfied at $(0, 0)$, and there exists a connected region G containing the origin in which they are satisfied.

The tangents to C_1, C_2 are $t_1: (x_u), t_2: (y_v)$. Let θ be the angle from t_1 to t_2 . Then $(x_u y_v) = \cos \theta$ and $(x_u y_v)^* = \sin \theta$. Furthermore, we find from (2) that the vectors n_1, n_2 with the direction cosines $(x_{uu})/(x_{uu} x_{uu})^{1/2}, (y_{vv})/(y_{vv} y_{vv})^{1/2}$ are normal to C_1, C_2 .

LEMMA 1. *In the region G*

$$\frac{(x_{uu} y_v)}{(x_{uu} x_{uu})^{1/2}} = - \frac{(x_u y_{vv})}{(y_{vv} y_{vv})^{1/2}} = (x_u y_v)^* = \sin \theta.$$

The ξ_2 -axis is directed downward, but we have agreed to consider the rotation positive which carries the positive ξ_1 -axis into the positive ξ_2 -axis. Then at $(0, 0)$ we find that the angles from t_1 to n_1 and t_2 to n_2 are $\pi/2$. Then because of (2) and (14) these angles are $\pi/2$ throughout G . Then the angle from n_1 to t_2 is $(\theta - \pi/2)$, and $(x_{uu} y_v)/(x_{uu} x_{uu})^{1/2} = \cos(\theta - \pi/2)$. Also the angle from t_1 to n_2 is $(\theta + \pi/2)$, and $(x_u y_{vv})/(y_{vv} y_{vv})^{1/2} = \cos(\theta + \pi/2)$. The lemma follows.

Lemma 1 enables us to write the equations of motion (5), (13) in G in the form

$$(18) \quad \begin{aligned} u'' &= \frac{m \sin \theta [(x_{uu}x_{uu})^{\frac{1}{2}} \cos \theta u'^2 + (y_{vv}y_{vv})^{\frac{1}{2}} v'^2] - g[m \cos \theta y_v^{(2)} - x_u^{(2)}]}{1 - m \cos^2 \theta}, \\ v'' &= \frac{-[(x_{uu}x_{uu})^{\frac{1}{2}} u'^2 + m(y_{vv}y_{vv})^{\frac{1}{2}} \cos \theta v'^2 - g x_u^{(1)}] \sin \theta}{1 - m \cos^2 \theta}. \end{aligned}$$

LEMMA 2. *In the region G*

$$\frac{(x_{uu}y_v)^*}{(x_{uu}x_{uu})^{\frac{1}{2}}} = \frac{(y_{vv}x_u)^*}{(y_{vv}y_{vv})^{\frac{1}{2}}} = -\cos \theta.$$

This lemma is proved by straightforward methods, using the facts stated in the proof of Lemma 1 about the angles between n_i and t_i .

LEMMA 3. *The part of the curve $\sin \theta \equiv (x_u y_v)^* = 0$ which lies in G consists of a single branch whose slope is everywhere positive, and which passes through the origin and divides G into two parts.*

The fact that the curve $\sin \theta = 0$ passes through the origin follows from (7). The slope of $\sin \theta = 0$ is $dv/du = -(x_{uu}y_v)^*/(x_u y_{vv})^*$, and by Lemma 2 this is

$$(19) \quad \frac{dv}{du} = \frac{(x_{uu}x_{uu})^{\frac{1}{2}}}{(y_{vv}y_{vv})^{\frac{1}{2}}};$$

hence, in G the slope is always positive. Because of (15) and (16) the curve crosses the v -axis at a single point only in G , and this crossing is at the origin. The curve is analytic at every point of G and has no multiple points. The single branch through the origin therefore extends into the first and third quadrants and continues with a positive slope until it passes out of G .

THEOREM 1. *If h is chosen sufficiently small so that the region of motion R is homeomorphic to a circular disc and lies in G , every orbit crosses the curve $\sin \theta = 0$ an infinite number of times.*

First, we notice from (17) that $-\pi/2 < \theta < \pi/2$ in G . Furthermore, $\sin \theta$ is positive in G above the curve $\sin \theta = 0$ and negative below it. Then since (15) holds, and since R lies in G , we see from (18) that $v'' = N \sin \theta$, where N is a function which is negative at all points of R . The theorem follows.

Assume henceforth that h is chosen so that the hypotheses of Theorem 1 are satisfied. Then R lies in G , contains the origin O , and is bounded by a

simple closed curve Z . Let X be the point in the third quadrant at which $\sin \theta = 0$ crosses Z , and Y the point at which the positive v -axis crosses Z . Let the region of R defined by $u \leq 0$, $\sin \theta \geq 0$ be designated by S . Then S is bounded by the segment OY of the v -axis, the arc XY of Z , and the arc OX of $\sin \theta = 0$.

LEMMA 4. *In the region S*

$$-g[m \cos \theta y_v^{(2)} - x_u^{(2)}] \geq 0,$$

and the equality holds only at the origin.

From (7) and (16) we see that $x_u^{(2)} > 0$ for $u < 0$ and $y_v^{(2)} < 0$ for $v > 0$. By (17), $\cos \theta > 0$ everywhere in G . Then the desired inequality holds at all points of S in the second quadrant including the positive v -axis and the negative u -axis except for the origin, at which the equality holds by (7). It remains therefore to establish the inequality in the part of S in the third quadrant.

Now since $y_v^{(2)} > 0$ for $v < 0$, and since $0 < m < 1$, we have

$$-[my_v^{(2)} \cos \theta - x_u^{(2)}] > -[y_v^{(2)} \cos \theta - x_u^{(2)}]$$

for $v < 0$. Now by using (2) we find that

$$[y_v^{(2)} \cos \theta - x_u^{(2)}] = y_v^{(1)} \sin \theta.$$

Then since $y_v^{(1)} < 0$ in G by (15), and since $\sin \theta > 0$ in S except on the boundary OX , where it vanishes, the desired inequality follows.

From Lemma 4 and (18) we see that in S we can write $u'' = M$, where $M \geq 0$, the equality holding only at the origin. In S , therefore, the equations of motion take the form

$$(20) \quad \begin{aligned} u'' &= M, \\ v'' &= N \sin \theta. \end{aligned}$$

THEOREM 2. *If h is so chosen that R is homeomorphic to a circular disc and lies in G , there exists at least one periodic orbit which joins two points of Z ; the two points of Z lie in the second and fourth quadrants respectively, and the orbit crosses $\sin \theta = 0$ at the origin and only at this point.*

Consider the orbit which at time $t = 0$ passes through an interior point P of the arc XY of Z . As long as it remains in S , the equations of motion can be written in the form (20). Then

$$u' = \int_0^t M dt, \quad v' = \int_0^t N \sin \theta dt,$$

and $u' > 0$, $v' < 0$ for $t > 0$ and thereafter until the orbit has passed out of S across XOY at Q . The slope of this orbit is therefore negative in S except possibly at P , and it is surely negative at Q . Now by Lemma 3 the slope of XO is everywhere positive; hence, the orbit is not tangent to XOY at Q , and Q varies analytically with P .

Let r be a coördinate which specifies the position of P on Z , r varying from 0 to 1 as P varies from X to Y . Also let s be a coördinate which specifies the position of Q on XOY , s varying from 0 to 1 as Q varies from X to Q , and from 1 to 2 as Q varies from O to Y . Then as shown above, $s = f(r)$, where $f(r)$ is analytic for $0 < r < 1$. Since the slope of the orbit is everywhere negative in S except possibly at P , we find $\lim s = 0, 2$ as r approaches 0, 1 respectively. Then since the curve $s = f(r)$, $0 < r < 1$, is continuous, it has at least one intersection with the line $s = 1$.

We have thus shown that there is at least one point P , say P_1 , for which the corresponding Q is at the origin O . Now the system is symmetric in the origin as specified by (12) by hypothesis; hence, the image in the origin of the orbit which joins P_1 to O is also an orbit OP_2 [5, section 2, Theorem 4], and the continuation of P_1O . Then P_1OP_2 is a periodic orbit which joins two points of Z [5, section 2, Theorem 3] and crosses $\sin \theta = 0$ at the origin and only at this point. Since the slope of P_1O is everywhere negative except possibly at P_1 , we see that P_1 is in the second quadrant and P_2 in the fourth. Also the orbit is without double points. The proof of the theorem is complete.

The theorem which we have just proved does not show that the periodic orbits belong to a group of periodic orbits obtained by analytic continuation with μ (which is essentially the energy constant) as the parameter. Certain results in the large on the analytic continuation of periodic orbits are known [7, 8, 9], but they do not suffice to prove the existence of a periodic orbit with unrestricted variation of μ .

THE UNIVERSITY OF ROCHESTER,
ROCHESTER, NEW YORK.

BIBLIOGRAPHY.

1. Birkhoff, "Dynamical systems with two degrees of freedom," *Transactions of the American Mathematical Society*, vol. 18 (1917), pp. 199-300.
2. Birkhoff, "Dynamical systems," *American Mathematical Society Colloquium Publications*, vol. 9 (1927).

3. Cederberg, "On the solutions of the differential equations of motion of a double pendulum," *Augustana Library Publications*, No. 9 (1923), Augustana Book Concern, Rock Island, Illinois.

4. Morse, "The foundations of a theory in the calculus of variations in the large," *Transactions of the American Mathematical Society*, vol. 30 (1928), pp. 213-274.

5. Price, "On reversible dynamical systems," *Transactions of the American Mathematical Society*, vol. 37 (1935), pp. 51-79.

6. Whittaker, *Analytical Dynamics*, 3rd edition (1927), Cambridge University Press.

7. Poincaré, *Méthodes Nouvelles de la Mécanique Céleste*, vol. 1 (1892), vol. 3 (1899).

8. Wintner, "Beweis des E. Strömgren'schen dynamischen Abschlussprinzips der periodischen Bahngruppen im restringierten Dreikörperproblem," *Mathematische Zeitschrift*, vol. 34 (1931), pp. 321-349.

9. Price, "On the Strömgren-Wintner natural termination principle," *American Journal of Mathematics*, vol. 55 (1933), pp. 303-308.

SPACE CURVES BELONGING TO A NON-SPECIAL LINEAR LINE COMPLEX.

By C. R. WYLIE, JR.

Introduction. By a "curve belonging to a linear line complex" is meant a curve all of the tangents to which belong to a linear line complex. Such curves exist, and have been the subject of a number of papers.

In his Doctor's Dissertation in 1877, Picard¹ cited an example of a sextic of genus one which belonged to a linear complex. In 1892, however, Steinmetz² erroneously concluded that curves of a complex had to be rational and moreover could have no cusps. Then in 1907, Snyder³ demonstrated a 1:1 correspondence between curves belonging to a linear complex and curves lying on a quadric cone. Since curves of every genus can be found on a quadric cone this result confirmed Picard and refuted Steinmetz. Finally in 1925, Loria⁴ reproduced the argument and conclusions of Steinmetz without perceiving the error.

From this point the present paper proceeds to an alternative development of the mapping relation between curves of a complex and curves on a quadric cone, derives the general parametric equations of such curves, deduces certain of their metrical properties, cites examples of curves of every genus which belong to a complex, proves that such curves may have any preassigned self-dual singularity, studies the $n-3 : n-3$ correspondence on complex curves with particular reference to a certain pair of sextics, and treats of the two families of rational sextics having six inflexions, those which belong to a linear complex and those which do not.

In order to find the parametric equations of complex curves it is necessary to have equations whereby the point common to two consecutive intersecting line can be found when the Plücker coördinates of the lines are given as functions of a single parameter. These equations are

¹ E. Picard, "Application de la Théorie des Complexes Linéaires à l'étude des Surfaces et des Courbes Gauches," *Annales de l'École Normale*, 1877.

² C. P. Steinmetz, "Curves which are self-reciprocal in a linear nul-system," *American Journal of Mathematics*, vol. 14 (1892), pp. 161-185.

³ V. Snyder, "Twisted curves whose tangents belong to a linear complex," *American Journal of Mathematics*, vol. 29 (1907), pp. 279-288.

⁴ G. Loria, *Curve Sghembe Speciali*, vol. 2 (1925), p. 21.

$$\begin{aligned}
 & \rho x_1 = P_{12}P'_{13} - P_{13}P'_{12} \\
 & \rho x_2 = P_{12}P'_{23} - P_{23}P'_{12} \\
 & \rho x_3 = P_{13}P'_{23} - P_{23}P'_{13} \\
 & \rho x_4 = P_{14}P'_{23} + P_{42}P'_{13} + P_{34}P'_{12}
 \end{aligned}
 \quad \text{where } P_{ik} = P_{ik}(t), P'_{ik} = \frac{dP_{ik}(t)}{dt}$$

1. It is well known from line geometry that there is a 1:1 correspondence between the lines of S_3 and the points of a hyperquadric V_4^2 of S_5 . Moreover the surface of tangents to a curve in S_3 is represented in S_5 by a curve which lies with its tangents on V_4^2 . If the curve in S_3 belongs to a linear complex which is not special its image in S_5 lies with its tangents in the section of V_4 by a hyperplane S_4 which is not tangent to V_4^2 . If in such a hyperplane the image curve is projected from a point on the V_3^2 , intersection of V_4^2 and S_4 onto an S_3 there is obtained a space curve whose tangents all meet a fixed conic, and whose osculating planes therefore are all tangent to the same conic. This conic is the image in the projection of the quadric cone cut from V_4 by the tangent S_3 at the center of projection. By duality in space this configuration is sent into a curve lying on a quadric cone.

To find the equations of curves of a complex let

$$V_4^2 \equiv P_{12}P_{34} + P_{13}P_{42} + P_{14}P_{23} = 0$$

be the Plücker identity and

$$S_4 \equiv P_{13} + 2P_{42} = 0$$

be the linear complex containing the original curve. Then

$$V_3^2 \equiv P_{12}P_{34} - \frac{1}{2}P_{13}^2 + P_{14}P_{23} = 0$$

and if the center of projection is

$$P_{23} = 1, \quad P_{12} = P_{13} = P_{14} = P_{34} = 0$$

and

$$P_{23} = 0$$

the S_3 of the projection, the conic, image of the cone of singular elements is

$$\begin{cases} 2P_{12}P_{34} - P_{13}^2 = 0 \\ P_{14} = 0 \end{cases} \quad \text{or} \quad P_{12} : P_{13} : P_{34} : P_{14} = \frac{1}{2}t^2 : t : 1 : 0.$$

Then

$$\frac{1}{2}t^2P_{34} - tP_{13} + P_{12} - f(t)P_{14} = 0$$

represents a family of planes all of which are tangent to this conic. The parametric point equations of the cuspidal edge of this family of planes are the Plücker coordinates of the tangents to the original curve. These are

$$P_{12} = f - t f' + \frac{1}{2} t^2 f''$$

$$P_{13} = -f' + t f''$$

$$P_{34} = f''$$

$$P_{14} = 1$$

$$P_{42} = \frac{-f' + t f''}{-2}$$

$$P_{23} = -f f'' + \frac{1}{2} f'^2$$

whence from equations A the parametric point equations of the original curve are

$$B: \begin{aligned} x_1 &= -t \\ x_2 &= f - \frac{1}{2} t f' \\ x_3 &= -f' \\ x_4 &= 1. \end{aligned}$$

If f is an implicit function of t so that $\phi(f, t) = 0$ these equations become

$$C^5: \begin{aligned} x_1 &= -t \partial \phi / \partial f \\ x_2 &= f \partial \phi / \partial f + \frac{1}{2} t \partial \phi / \partial t \\ x_3 &= \partial \phi / \partial t \\ x_4 &= \partial \phi / \partial f. \end{aligned}$$

In each case the genus of the curve is the genus of the function connecting f and t .

2. The parametric plane equations of these curves each contain the factor f'' , hence $f''' = 0$ is the condition for an inflexion. The condition for stationary planes reduces to $f''' = 0$. Hence, *The stationary planes of a curve belonging to a linear complex coincide in pairs at the inflexions.*

If we put

$$\begin{aligned} x &= -t \\ y &= f - \frac{1}{2} t f' \\ z &= -f' \end{aligned}$$

so that we have the metric case, we find by elementary calculations that the curvature of a complex curve is

$$\frac{1}{\rho} = \frac{f'''}{2(s')^3} \sqrt{t^2 + 4 + f'^2}$$

* B. Segre, "Sulle Curve le cui Tangenti Appartengono al Massimo numero di Complessi Lineari Indipendenti," *Memorie dell' Accademia dei Lincei*, series 6, vol. 2 (1928), pp. 578-592.

s , the arc length parameter. The torsion reduces to the simple form

$$\frac{1}{\sigma} = \frac{2}{t^2 + 4 + f'^2}.$$

Thus, *At every real finite point the torsion $1/\sigma$ of a curve belonging to a linear complex is*

$$0 < 1/\sigma \leq \frac{1}{2}.$$

From this formula it is easy to deduce that the only curves of constant torsion belonging to a linear complex are helices on circular cylinders having OY as axis.

3. If in equations C , we put $\phi \equiv f^{2p+1} - at^{p+1} - bt^p = 0$ we have the curves

$$\begin{aligned} x_1 &= -(2p+1)t(at^{p+1} + bt^p)^{[2p/(2p+1)]} \\ x_2 &= \frac{1}{2}a(3p+1)t^{p+1} + \frac{1}{2}b(3p+2)t^p \\ x_3 &= -a(p+1)t^p - bpt^{p-1} \\ x_4 &= (2p+1)(at^{p+1} + bt^p)^{[2p/(2p+1)]}. \end{aligned}$$

These are of order $5p+2$ and of genus p . Each has a point of type $(p, p+1, p)$ at $(1, 0, 0, 0)$ and a point of type $(p+1, p, p+1)$ at $(0, 0, 1, 0)$. They lie on the ruled cubic surface

$$x_1(2bx_3x_4 - ax_1x_3 + ax_2x_4) - p(ax_1 - bx_4)(3x_1x_3 - 2x_2x_4) = 0.$$

If more generally we put $\phi \equiv bf^{KN} - at^{LN} - 1 = 0$ where $L > K$ and L is prime to K , we have the curves

$$\begin{aligned} x_1 &= -Kbtf^{KN-1} \\ x_2 &= \frac{1}{2}a(2K-L)t^{LN} + K \\ x_3 &= -Lat^{LN-1} \\ x_4 &= Kbf^{KN-1}. \end{aligned}$$

These are of genus $\frac{(LN-1)(KN-1) - (N-1)}{2}$ and of order KLN^2 except when $K=1$ and $L=2$, in which case the order becomes $2N^2 - N$. These curves have a singularity of type $[(L-K)N, (2K-L)N, (L-K)N]$ at $(0, 1, 0, 0)$ when $L < 2K$; they are non-singular when $L = 2K$, i. e. when $K=1, L=2$; and they have a singularity of type $[KN, (L-2K), KN]$ at $(0, 1, 0, 0)$ when $L > 2K$. For proper choices of K, L, N , these singularities can be made to represent any arbitrary self-dual singularity, say (s, t, s) . Hence,

* M. F. Egan, "The linear complex and a certain class of twisted curves," *Proceedings of the Royal Irish Academy*, vol. 29 (1911), pp. 33-72.

Curves can be found belonging to a linear complex and having any preassigned self-dual singularity counted any number of times.

4. Curves belonging to a linear complex have an $n-3 : n-3$ correspondence on them generated by the residual intersections of the curve and its osculating planes, and the points of contact of the osculating planes drawn to the curve from its various points. This correspondence is always composite, having at least one bilinear factor.⁷

If in equations C , we choose ϕ to be either

$$f^2 - 4t^3 + c^2t = 0 \quad \text{or} \quad f^2 - 4t^3 + d^3 = 0$$

so that it is harmonic or equianharmonic respectively, the resulting curves are sextics of genus one, and the $3:3$ correspondence defined on them by

$$\sum_{i=1}^4 \frac{x_i(t)u_i(s)}{(t-s)^3} = 0 \quad \begin{array}{l} x, \text{ the parametric point equations} \\ u, \text{ the parametric plane equations} \end{array}$$

factors rationally into three bilinear factors. In each case these factors with the identity form a group of order four leaving the curve invariant. The points of each curve are thus arranged in tetrads such that the osculating plane at any one of them passes through the other three. The tetrads must consist then of collinear points, so that each curve has a regulus of quadri-secants, among which are three bitangents. These are the simplest irrational curves analogous to the rational quintic discussed by Egan.⁸

5. In 1877, Picard⁹ showed that if a rational curve belonged to a complex it had to have $2(n-3)$ inflexions. In 1901 Grace¹⁰ attempted to prove the sufficiency of this necessary condition. His argument was faulty however, and Egan¹¹ pointed out the error and gave an example of a sextic with six inflexions which did not belong to a linear complex. Gherardelli¹² elaborated this by showing that there were two classes of rational sextics, one of which consisted of curves belonging to a complex, the other of which consisted of

⁷ A. Terracini, "Sulla Riducibilit  di Alcune Particolari Corrispondenze Algebriche," *Rendiconti Circolo Matematico di Palermo*, vol. 56 (1923), pp. 112-143.

⁸ M. F. Egan, *loc. cit.*

⁹ E. Picard, *loc. cit.*

¹⁰ J. H. Grace, "Curves in a linear complex," *Proceedings of the Cambridge Philosophical Society*, vol. 11 (1901), pp. 132-134.

¹¹ M. F. Egan, *loc. cit.*

¹² G. Gherardelli, "Le Sestiche Sghembe Razionali con Sei Flessi," *Rendiconti dell' Accademia dei Lincei*, series 6, vol. 2 (1930), pp. 173-179.

curves not belonging to a complex. He derived the first family of curves projecting the rational normal sextic Γ_6 of S_6 onto an S_3 from a plane meeting six osculating planes of Γ_6 , which plane did not lie on the V_5^2 , fundamental for the polarity determined by Γ_6 . The second family he derived by projecting Γ_6 from a plane meeting six osculating planes of Γ_6 and lying on the fundamental hyperquadric V_5^2 . Curves of this second family he showed had their inflexions coplanar, and the osculating planes at the inflexions concurred.

In the present paper the first class of sextics is obtained by sectioning the V_3^{12} of osculating planes to Γ_6 with an S_4 through two osculating rays of Γ_6 which are not coplanar, and the second class is obtained by sectioning the V_2 of tangents to Γ_6 with an S_5 in general position, in each case interpreting the curve of intersection as a configuration on the fundamental hyperquadric V_4 of line geometry. These results will not be developed here.

CORNELL UNIVERSITY.